



[illegible]

QY	912	ACTACCTTAGTAAAGTATTAATAAGAAATGTGATATCATCATTAATAATATATG--	-CCT	969
Db	929	ATTATATTATTATTATTATTATTAATAATATTTATTATTAATTTATNTATTAATAATATATTATATAT		988
QY	970	TATTTAAATTAATGATAAGTTGTATCATTTAAAGTTGAAAAACCAATAGCTCCGCT		102
Db	989	TATTAATAATAATTTATTTATATATTATTAATAATAATAATAATTAAGTTTATATTATTAATATTA		104
QY	1030	TAATTTTGAATATATGTTTCTATGTTACTTTCTCAAGCCATATATAAAAACTTTGTA		1088

```
Db 1049 TAAATTAATTAACGATTAATAATTAATTAATTAATTAATTAACGATTAATTAATTTAG 1108
Qy 1090 ATGCTAAATTTGATGCTGGAAGAAAAATGCTGAATGCAATGAAATTAATGCTATTC 1149
Db 1109 ATTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1168
Qy 1150 AAAGTCCAAAATCCATCAATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1209
Db 1169 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1228
Qy 1210 TTTTAAATTTTACAGCAATTAATAATTTCTATTTTAAATTTTACATATTAAT 1269
Db 1229 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1288
Qy 1270 TTAATCACTGTCACTTTAGAAATACCAACAATTAATTAATTAATTAATTTATCTT 1329
Db 1289 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1348
Qy 1330 AATAATTTGAGATCTCTCAATTAATCTGATTAATTTTAATTTTGTGATATTTCT 1389
Db 1349 ATATTAATTTGTTGGTTGGAATTAATTTGGGTTGTTTTTTTTTTTTTTTTTTTGA 1408
Qy 1390 TATGTTTGAAGTT 1403
Db 1409 GATGAGGTTTGT 1422
```

## RESULT 4

```
US-10-240-485-46/c
; Sequence 46, Application US/10240485
; Publication No. US20030148327A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with
; FILE REFERENCE: 5013.1007
; CURRENT APPLICATION NUMBER: US/10/240,485
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03970
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 202
; SEQ ID NO 46
; LENGTH: 11996
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-240-485-46
```

```
Query Match 4.4%; Score 77.8; DB 12; Length 11996;
Best Local Similarity 50.1%; Pred. No. 0.0011;
Matches 248; Conservative 0; Mismatches 242; Indels 5; Gaps 2;
```

```
Qy 912 ACTACCTTAGTAAGATATATAGAAATTTAATCATGACATTTAATTTATGT--CCT 969
Db 11068 ATTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11009
Qy 970 TATTTTAAATTTGATTAAGATTGATCAATTAAGATTGAGAAAACCAATAGCTCGTCT 1029
Db 11008 TATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10949
Qy 1030 TGAATTTGAAATTAATTTGTTTCTATGTAATCTTTCTCAAGCTATATAAAAACTTTGTA 1089
```

```
Db 10948 TATATTAATTAACGATTAATTAATTAATTAATTAATTAATTAATTAATTAACGATTAATTAATCTTAA 10889
Qy 1099 ATGCTAAATTTGATGCTGGAAGAAAAATGCTGAATGCAATGAAATTAATGCTATTC 1149
Db 10888 ATTATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10829
Qy 1150 AAAGTCCAAAATCCATCAATGAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1209
Db 10828 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10772
Qy 1210 TTTTAAATTTTACAGCAATTAATAATTTCTATTTTAAATTTTACATATTAAT 1269
Db 10771 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10712
Qy 1270 TTAATCACTGTCACTTTAGAAATACCAACAATTAATTAATTAATTAATTTATCTT 1329
Db 10711 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10652
Qy 1330 AATAATTTGAGATCTCTCAATTAATCTGATTAATTTTAATTTTGTGATATTTCT 1389
Db 10651 ATATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10592
Qy 1390 TATGTTTGAAGTTA 1404
Db 10591 TAAATTAATAATCTTA 10577
```

## RESULT 5

```
US-10-311-455-288
; Sequence 288, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 288
; LENGTH: 6641
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-288
```

```
Query Match 4.3%; Score 76.4; DB 12; Length 6641;
Best Local Similarity 45.9%; Pred. No. 0.0015;
Matches 372; Conservative 0; Mismatches 431; Indels 7; Gaps 3;
```

```
Qy 755 AATATGTCAGTTACCTTCTGCAATTAAGAGCACTTAAGACACTTTTCTAAAT 814
Db 2561 ATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2620
Qy 815 ACTTCAAGGAGGAGATTTTAACTTAAGAGGATTAAGAGGAGGAGGAGGAGGAGGAGGAG 874
Db 2621 ATATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2680
Qy 875 TAAAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 934
Db 2681 TGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2740
Qy 935 GAAATTTGTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 994
```

```

Db      2741 TTAATATAGAGATGTTTATATTAATATATTTTAAAGAAATATAGTAGATAT 2800
      995 TCATTAGATTGAGAAACCAATAGCTCGTCTGTGATTTTGAATTTATGTTCTAT 1054
      2801 TGAATTAATTTTGGTTAGAGATATGATTTATTTTATTTAGATTTTGAAGTTATAT 2860
Qy      1055 GTTACTTTCTTCAAGCCTATATATAAACTTTGTA--ATGCTAAATGTAAGCTGAAA 1111
      2861 GTTAATTAATGAATTTATTTATTTAGAAAGATATAGAGATTAATAAATAGATATAAA 2920
Qy      1112 AAAATGTGAATGAATTCATATAGAAATATG--TATTTCAAGTCCAAATGCATCAAT 1169
      2921 GTTATATTTAAATTAATGATATTTTATTTATTTTGTGATTTAAATTAATGATTTAT 2980
Db      1170 AGAATTTAGTACAAAGTACCAAAATATCTCTATTTTAAATTTTACACAAAT 1229
      2981 AATAAGTATAGATTAATTTATTTATTTATTTAGAGATTAATTTATTTAATAATATA 3040
Qy      1230 TAAAAATATCTCTATTTTAAATTTTACAAATATATATATATCATCTGTCACTTTAG 1289
      3041 GTTAATTTATTTATTTATATGATTTAATTTAGTATTTATTTATTTATTTATTTATTA 3100
Qy      1290 AATACCACCAACATATTAATTAATCTAGATTTATTTATTTCTTAATTTTGAATCTCTCA 1349
      3101 ATTAGTAAATTTATTTATTTATTTATTTAGATATATAAATAGTAATTTATTTATTTAT 3160
Qy      1350 ATATATCTGATTTATTTATTTA--TATTTGTCATATTTCTTATGTTTATAGTAGTAA 1407
      3161 AATAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3220
Qy      1408 CTATATCTTGTGCAACTAGTATTTCAATATATGAGTTTGTGAGACACATTTGACATC 1467
      3221 AATAATTTTATTTTAACTATTTTAAATAGTAAATTTGATTTTAAATTAATTA 3280
Qy      1468 TTGAACATTTGTTTAACTTTGTGGAGTAAAGGTAAAGCAATTAACATTCAGATTAAG 1527
      3281 GTAGAAATTTATTTATTTATTTATTTATTTATTTATTTATTTAATAATAGTTGTTTAA 3340
Qy      1528 ACCATCTATTAATATCTCTTGTGCTTT 1557
      3341 TATATGATTTAGGCTTTAAATTTTATTTT 3370
Db

```

## RESULT 6

```

US-10-240-452-36
; Sequence 36, Application US/10240452
; Publication No. US20030162194A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIERENBERG, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Apoptosis
; FILE REFERENCE: 5013.1006
; CURRENT APPLICATION NUMBER: US/10/240.452
; CURRENT FILING DATE: 2002-10-02
; PRIOR APPLICATION NUMBER: PCT/EP01/03969
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: DE 10019058.8
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: DE 10019173.8
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 78
; SEQ ID NO 36
; LENGTH: 6641
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURES:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)

```

## US-10-240-452-36

Query Match 4.3%; Score 76.4; DB 12; Length 6641;  
 Best Local Similarity 45.9%; Pred. No. 0.0015;  
 Matches 372; Conservative 0; Mismatches 431; Indels 7; Gaps 3;

```

Qy      755 AATGTGAGTTACCTTTCTGCACTTATAGAGCACTTACGACACTTTACTTAAT 814
      2561 ATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 2620
Db      815 ACTCAAGAGAGAGATTTTAACTTATAGAGATTAAGGAGTTAAAGACACACAT 874
      2621 ATATTAAGTAATTAATTTATTTATTAATTAAGATATATGATTTAAGTAGTTGTATG 2680
Qy      875 TAAAGGAGCTTTAAATTAATGTTGTGAACACACACTTCTTATGATTTATTA 934
      2681 TGAAGAGTTAGATTTTATTTTATTTTGTATTTTAAAGATTTATTTATTTATTTAT 2740
Qy      935 GAAATGTATCATCATATATATATTTTGTCTTATTTAAATTTATGATTAAGTTGA 994
      2741 TTAATATTAAGATTTTATTTATTAATAATTTTAAAGAAATATGTTAGATAT 2800
Qy      995 TCATTAGATTGAGAAACCAATAGCTCCGCTGTGATTTTGAATTTTGTTTTCTAT 1054
      2801 TGAATTAATTTTGTGTTTACAGATTTAGATTTATTTTATTTTATTTTATTTAT 2860
Qy      1055 GTTACTTTCTTCAAGCCTATATATAAACTTTGTA--ATGCTAAATGTAAGCTGAAA 1111
      2861 GTTATATTAATTAATTTATTTATTTAGAGATATATGAGATTAATAATTAAGATATAAA 2920
Qy      1112 AAAATGTGAATGATTTCAATAGAAATTTAG--TATTTCAAGTCCAAATGCATCAAT 1169
      2921 GTTATATTTAAATGAATTTTATTTTATTTTGTGATTTTAAATTAAGATTTAT 2980
Qy      1170 AGAATTTAGTACAAAGTACCAAAATATCTCTATTTTAAATTTTACACAAAT 1229
      2981 AATAAGTATAGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3040
Qy      1230 TAAAAATATCTCTATTTTAAATTTTACAAATATATATATATATCATCTGTCACTTTAG 1289
      3041 GTTAATTTATTTATTTAAAGATTTATTTAGTAATTTATTTATTTATTTATTTATTA 3100
Qy      1290 AATACCACCAACATATTAATTAATCTAGATTTTATTTATTTCTTAATTTTGAATCTCTCA 1349
      3101 ATTAGTAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3160
Qy      1350 ATATATCTGATTTATTTT--TATTTGTCATATTTCTTATGTTTATAGTAGTAA 1407
      3161 AATAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3220
Qy      1408 CTATATCTTGTGCAACTAGTATTTCAATATATGAGTTTGTGAGACACATTTGACATC 1467
      3221 AATAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3280
Qy      1468 TTGAACATTTGTTTAACTTTGTGGAGTAAAGGTAAAGCAATTAACATTCAGATTAAG 1527
      3281 GTAGAAATTTATTTATTTATTTATTTATTTAAGTATTTAATAATTAAGGTTGTTTAA 3340
Qy      1528 ACCATCTATTAATATCTCTTGTGCTTT 1557
      3341 TATATGATTTAGGCTTTAAATTTTATTTT 3370
Db

```

## RESULT 7

```

US-10-311-455-227
; Sequence 227, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIERENBERG, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OR INVENTION: cytosine methylation

```



```
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 227
; LENGTH: 18154
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-227
```

```
Query Match 4.3%; Score 76.4; DB 12; Length 18154;
Best Local Similarity 4.2%; Pred. No. 0.0022;
Matches 303; Conservative 0; Mismatches 331; Indels 8; Gaps 2;
```

```
QY 927 TATATAGAGAAATGTATCATCATCATTAATTAATTCCTTATTTAAATTAAGATA 986
Db 13819 TTTTATTAATTAATTTGATATATTAATTTTAAATTAATTTGCTTTT 13878
QY 987 AAGTGTATCATTAAGATGAGAAACCAATAGTCTGCTGATTTTGAATTAATG 1046
Db 13879 TAAATTTATTAATGA-ATAATTTAGTATTTGATTAATTAATTAATTAAT 13937
QY 1047 TTTTCATGTACTTCTTCTCAAGCCATATATAAATTTGTAATGCTAATGCT 1106
Db 13938 TATTTTATTTATTTTATTAATTAATTTGATTAATTAATTTTAAATGCTTAA 13997
QY 1107 GGAAGAAATGTATGAATGAAATCAATGAATTAATGTAATTCAGATCCAAATCCATC 1166
Db 13998 TGATTTGATGTTTGTATTTTAAATTAATTTTGAATGATTAATTTTATTTATG 14057
QY 1167 AATAGAAATTTAGTACAAAACGTACTCAAAAATTTCTCTTATTTAAATTTTACACA 1226
Db 14058 TTTTATTTTATTTTGTTTTATTTTAAATTTTGAATTTTGAAGTTTGTG----- 14112
QY 1227 ATATATAAATATCTCTTATTTTAAATTTTACATATATAATTAATTCACCTGTCACCT 1286
Db 14113 -ATTTGAATATATTTATTTATTTTAAATTTTAAATGTAATTTATTTTAA 14170
QY 1287 TAGAATACCAACAATATTAATTAATTAATTTTATTTTAAATTTTGAATCTC 1346
Db 14171 TAAATTTGAAAGTATTTTATTTAGTTAGTTAATTAATTTATTAATTAATTAAT 14236
QY 1347 TCAATATATCTGATTTATTTATTTATTTGTCATATTTCTTATGTTTGAAGTAAAC 1406
Db 14231 TTTTAAATTTTATTTTATTTTATTTGATTTTATTTTAAATTTATTTATTTGTAATGT 14290
QY 1407 CCTTATATCTGTGCAACTAGTAAATCAATATATATGAGTTTGAAGACACATTCACAT 1466
Db 14291 ATTTTAAAGTTGTATATATATATATTTATGATAGATATGCTGTGCTTTATTAAGTG 14350
QY 1467 CTGGAACAATGCTTTTAACTGTTGGAATGTTAAAGTAAATTAACCTTGAATTAAT 1526
Db 14351 TTTTATTTTATTTTATTTTATTTATTTATTTGTAATTAATTAATTTTGAAGATTAATTA 14410
QY 1527 GACCATCTATTAATATCTCTGCTTTGCTTTTAAATTAAGTGT 1568
Db 14411 TATATATAGGAATATATAGTTTATTTATTTGTAAGAAATTTT 14452
```

```
RESULT 8
US-10-311-455-2096/c
; Sequence 2096, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEX, Alexander
```

```
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Deter
; TITLE OF INVENTION: cytosine methylation
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311.455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2096
; LENGTH: 10329
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2096
```

```
Query Match 4.2%; Score 74.2; DB 12; Length 10329;
Best Local Similarity 4.78%; Pred. No. 0.0042;
Matches 313; Conservative 0; Mismatches 333; Indels 9; Gaps 3;
```

```
QY 712 TTAACATAAACAACCACTTCATTCGAGAGATGATGAAGAAAGAAATGCGAGTACCTT 771
Db 5982 TTCATTTTAAACCAATTAATGACAAATCTTATTAATTAATTTATTAATTAAC 5923
QY 772 TCTGAGTTCATAGAGCAACTTACAGACACTTTTACTAA---ATACTCMAAGAGA 827
Db 5922 TATTCATTCATTAATTTATATATATATATCCATTAATAAATCAATCCTCCTTATTTA 5863
QY 828 AGATTTTAAACCTAGAGAAATGATGGAGTTAAAGCAACACATTAAGGGAGAGT 887
Db 5862 AAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5803
QY 888 TAAATTTAATGTTGTATACCAACCACTACCTTTGTAATTAATTAAGAAATTTGTAATC 947
Db 5802 TTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5743
QY 948 ATCAATATATATTTATTTCTTATTTTAAATTTATGATTAAGTTGATCATTAAGATTGA 1007
Db 5742 AACTTATTTACACAAAATATTAATTTTAAATTAATTAATTAATTAATTAATTAATTA 5683
QY 1008 GAAAAACAATAGTCTGCTGCTGATTTTGAATTAATTTGTTCTATGTTACTTTCTTC 1067
Db 5682 TATAATTAATTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTA 5623
QY 1068 AAGCCTATATTAATTTGTTGTAAT-GCTAAATTTGATGCTGGAATAAATGTTGTAATGA 1126
Db 5622 AAGCAATTAATTAATTTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5563
QY 1127 TTCATTAAGAAATTTGATTTCAAGGCCAAATTCATATGAATTAAGTAAATTAAGTAA 1166
Db 5562 TTAATAAATTAATTTATTTCTGTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5503
QY 1187 CGTAACCAAAATATTTCTTATTTTAAATTTTAAATTTTAAATTAATTAATTAATTAAT 1246
Db 5502 CAATACCAATTAATTAATTTCCCAAAAAA---AAACAACTTAAATTAATTAATTAATTA 5447
QY 1247 TTTTAAATTTTAAATTAATTAATTTATTAACCTGTCACCTTTAGATTAACCAACAATAT 1306
Db 5446 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5387
QY 1307 TAAATCTAGATTTATTTATTTATTAATTAATTTGAGATCTGCATATATATGATTA 1361
Db 5386 ATATATTTTCCAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5332
```

```
RESULT 9
US-10-311-455-986
```

```

; Sequence 386, Application US/10311455
; Publication No. US20030143606a1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEBENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO: 986
; LENGTH: 7597
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
US-10-311-455-986

```

```

Query Match      4.2%; Score 74; DB 12; Length 7597;
Best Local Similarity 43.8%; Pred. No. 0.0041;
Matches 367; Conservative 0; Mismatches 470; Indels 1; Gaps 1;

```

```

QY 998 AATGATCATCATATATATATGCTTATTAATAATATAGTAAAGTGTATCA 997
DB 6 AAGGTAATGAGATATATATATGCTTATTAATAATATAGTAAAGTGTATCA 65
QY 998 TTAAGATGAGAAAACCAATATGCTGCTGATTTGAATTAATGTTTCTATGT 1057
DB 66 TTTGGGCTTATTTATAGTATATATGTTTATTTATTTATTTATTTATTTATTTATG 125
QY 1058 ACTTTCTCAAGCCATATATAAAGCTTGTAAGCTAAATTTGATGCGGAAGAAAG 1117
DB 126 TATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 185
QY 1118 TGTATGATTCATATAGAAATTAATGATTTCAAGTCCAAATTCATCATAGAAATTT 1177
DB 186 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 245
QY 1178 AGTCAAAAGCTATCTAAAATATCTCTATTTTAAATTTTACAAATATATAAATA 1237
DB 246 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 305
QY 1238 TTTCTTATTTTAAATTTTCAATATATATTTATTTATTTATTTATTTATTTATTTAT 1297
DB 306 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 365
QY 1298 CAACAATTTATCTAGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1357
DB 366 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 425
QY 1358 GATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1417
DB 426 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 485
QY 1418 GGTCAAACTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1477
DB 486 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 545
QY 1478 GGTATTAACCTTTGGAATGTTAAAGTAAATAAACATTCAGAAATATGACCATATTT 1537
DB 546 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 605
QY 1538 AATATCTCTCTTCTTTTAAAGTGTGATGAAATGCTCTATGTAAGCTAGAG 1597
DB 646 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 664

```

```

QY 1598 TGTCTTGCGCTGTGATATTCATTTCCAGATGTAAGAACTGCCACTACGA 1657
DB 665 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 724
QY 1658 TATTTAGCATTAACACAGTATGTTTAAACAGCTCCCTGATTTTGGCATATAT 1717
DB 725 TTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 784
QY 1718 TCGCTCTCTTCTTTTCTTCAAGTATTAACATGAACTATTTATAGAGCATGA 1775
DB 785 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 842

```

```

RESULT 10
US-10-312-841-1
; Sequence 1, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHO
; FILE REFERENCE: E01/1208/WO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO: 1
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure
; LOCATION: (3294164)
US-10-312-841-1

```

```

Query Match      4.1%; Score 73.6; DB 12; Length 3673778;
Best Local Similarity 44.3%; Pred. No. 0.046;
Matches 298; Conservative 0; Mismatches 374; Indels 0; Gaps 0;

```

```

QY 918 TTTAGTATGATTTATTAAGAAATTTGATATCATCATTTATTTATTTGCTTATTTAA 977
DB 1713988 TTTATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1714047
QY 978 ATATGATTAAGTGTATCATTTAGATGAGAAACCAATATGCTGCTGATTTT 1037
DB 1714048 TTTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1714107
QY 1038 GAATTTATTTCTATGTTTACTTTTCTTCAAGCCATTTATTAATAAATTTGTAATGCTTAA 1097
DB 1714108 TATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1714167
QY 1098 TGTATGCTGGAAGAAATGTAATGATTCATATGAAATTAATGATTTCAAGTCA 1157
DB 1714168 TTTATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1714227
QY 1158 AATTCATCAATGAAATTTAGTACAAAACGTAACCTCAAAAATTTCTTATTTTAAAT 1217
DB 1714228 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1714287
QY 1218 TTTCAACAATATTAATAATTTCTCTTATTTTAAATTTTACAAATATTTATTTATTTAT 1277
DB 1714288 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1714347
QY 1278 TGTACCTTTAGAAATCCACCAAAATTAATTAATCTTACATATTTATTTCTTAAATTT 1337
DB 1714348 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1714407
QY 1338 TGAGATCTCCAAATATATGATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1397
DB 1714408 TATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1714467
QY 1398 AGAGTTAACCTTATATCTTTGTCGAACATGATTAATCAATATATAGATTTGTAAGGACA 1457

```

RESULT 11  
US-10-240-485-146  
; Sequence 146, Application US/10240485  
; Publication No. US2003018327A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with  
; TITLE OF INVENTION: Metastasis  
; FILE REFERENCE: 5013.1007  
; CURRENT APPLICATION NUMBER: US/10/240 485

Query Match	4.18;	Score 72.6;	DB 12;	Length 6189;
Best Local Similarity	43.68;	Pred. No. 0.0061;		
Matches 488;	Conservative 0;	Mismatches 612;	Indels 18;	Gaps 3

QY	341	TGAAAATCTGGCGCAAGAGTTCTGAACTGTGCATATTTCTTAACAGTATAGAAAATTTCTGA	400
Db	2028	TGAATATAGATGTAAATAAGTTTGAAGANTATTTTTTTTTTGAGAGTCGTGGTGGTTTAA	208
QY	401	AGTGTTTAGAATTTTGACTTTTCCAAAGCAAACTTGACTTTTGACTTTCTTAATAAACA	460
Db	2088	GAATTTGGAAAGTAAGAGTATTTTATTAATAGATATAGATTTAGTTATTTTATTTATTAATAA	214
QY	461	AACTTCATATTTCTAACATGCTCTTGATGAAATGATGATCTTGAATTTGATGTGATGACA	520
Db	2148	AATATATGAATAAGAAATTAATGATAAATATTTATTTAATGTAAAGTTTAAATGGGTAAAT	220
QY	521	AAGTCAAAGTTTGACTTTTCAGTGTGCATTTGACCATTTTGGCTCTTGTGGCAATTTCCAA	580
Db	2208	TTTATTTGATATTAATTAATAGAGTTTGTATTTGTAATTAATTAAGATGTTTGGGAA	226
QY	581	CCATAATTGATGTATCACTGCTGCAAACTTGATGTCATGAGAGATCTTAATAGAAAATATTC	640
Db	2268	AAGTGTGTTTAAAGCGAATTTGTTTTTTTTTGTGGTTAAGATGTTTTTTTTTATGTTTTTTA	232
QY	641	TTGAAGACTGAGAGGAAAAATTTTGTAGTACACACAAAGAAATCCCTGTTTTTCATATGTCG	700

```

RESULT 12
US-10-240-453-100
: Sequence 100, Application US/10/240453
: Publication No. US20030148326A1
: GENERAL INFORMATION:
: APPLICANT: OLEK, Alexander
: APPLICANT: PIEPENBROCK, Christian
: APPLICANT: BERLIN, Kurt
: TITLE OF INVENTION: Diagnosis of Diseases Associated with DNA
: TITLE OF INVENTION: Transcription
: TITLE OF INVENTION: by Means of Assessing the Methylation Status of Genes Associated
: TITLE OF INVENTION: with DNA Transcription
: FILE REFERENCE: 5013.1009
: CURRENT APPLICATION NUMBER: US/10/240,453
: CURRENT FILING DATE: 2002-10-02
: PRIOR APPLICATION NUMBER: PCT/EP01/03973
: PRIOR FILING DATE: 2001-04-06

```

```

RESULT 13
US-10-239-676-90
; Sequence 90, Application US/10239676
; Publication No. US20030082509A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation
; FILE REFERENCE: 5013.1003
; CURRENT APPLICATION NUMBER: US/10/239,676
; CURRENT FILING DATE: 2002-09-24

```

```

RESULT 14
US-10-311-455-477/c
; Sequence 477, Application US/10311455
; Publication No. US20030143606a1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; APPLICANT: BERLIN, Kurt
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detecting
; TITLE OF INVENTION: Cytosine methylation

```

```
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311.455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 477
LENGTH: 6065
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-477

Query Match
Best Local Similarity 48.1%; Score 72.2; DB 12; Length 6065;
Matches 271; Conservative 0; Mismatches 283; Indels 9; Gaps 2;

QY 811 AAACTACAAAGAGAGATTTTAACTAGAGAGATGAGGATTAAGAGCAAC 870
DB 1247 AAAAAAAAAAATAAATCTTTAAATACAAACAAATCAATTAATAAATTTCC 1186
QY 871 ACATTAAGGGGAGTGTAAATTAATGTGTGTACACCACTACTAGTACTATT 930
DB 1187 ATAAATACTTAACCTTTAAATTAATATCTTTAAATACCA-----AAAAAAT 1136
QY 931 ATAAAGAAATTTGATCATCATATTAATTAATGCTTATTAATTAATGATTAAGT 990
DB 1135 ATACTAAACCTTAATCAAAATACCAATTTCTTTAACTTTAAACCAACAAAA 1076
QY 991 TGTATCATTAAGTGAAGAAACCAATAGTCTGCTGCTGATTTTGAATTAATGTTT 1050
DB 1075 ATTTTAAATATATTAATAAACAATAATTTTCAATCTT-TTCATAAATAATACTCT 1017
QY 1051 CTATGTACTTTCTTCAAGCTATATAAACTTTGTAATGTAATGTGCTGGA 1110
DB 1016 AATTTATATTTCAAAAAAATCCAAAAAATAAATACTAATAATAATAAATTAACCTTA 957
QY 1111 AAAAAATGTAATGAATTAATGAATTAATGTAATTTCAAGTCAAAATCCATCAATA 1170
DB 956 TAACACTTACAAATTAATTAATAAATAATTAATACTCAAAATTTTATATA 897
QY 1171 GAATTTAGTACAAACGTAACTCAAAATATTTCTTTTAAATTTTACACAAAT 1230
DB 896 AAAAAATTTACTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATCT 837
QY 1231 AAAAAATTTCTTTATTTTAAATTTTACATAATTAATTAATCAAGTCACTTTAGA 1290
DB 836 CCTCTATTTTAAATAATCATATATATTAATAAATCAATTAATAAATAAATAAATA 777
QY 1291 AATACCAACACATATTAATTAATTAATTAATTTTAAATTTTGAATTTGAGATCTCTCA 1350
DB 776 ATTTCTTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 717
QY 1351 TATATCTGATTTTATTTTAT 1373
DB 716 ACTAAATAACATAATATCT 694

RESULT 15
US-10-311-455-1128/c
Sequence 1128, Application US/10311455
Publication No. US20030143606A1
GENERAL INFORMATION:
APPLICANT: OLEX Alexander
APPLICANT: BERLIN, Kurt
APPLICANT: PIERENBERG, Christian
TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detect
TITLE OF INVENTION: cytosine methylation
```

```
FILE REFERENCE: 5013.1014
CURRENT APPLICATION NUMBER: US/10/311.455
CURRENT FILING DATE: 2002-12-16
PRIOR APPLICATION NUMBER: PCT/EP01/07537
PRIOR FILING DATE: 2001-07-02
PRIOR APPLICATION NUMBER: DE 10032529.7
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: DE 10043826.1
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 2424
SEQ ID NO 1128
LENGTH: 5807
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-1128

Query Match
Best Local Similarity 47.4%; Score 71.8; DB 12; Length 5807;
Matches 313; Conservative 0; Mismatches 342; Indels 6; Gaps 3;

QY 818 ACAAAGAGAGATTTTAACTAGAGAGATGAGGATTAAGGACACATTA 877
DB 5384 ATAACTAAATAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5325
QY 878 GGGGAGTGTAAATTAATGATGTAACCAACCACTTTAGTATTAAGAA 937
DB 5324 TAAACCAAAAAAATTAATCTCTAATAATAACCAATTAACCAATCTTTTATTTT 5265
QY 938 AATGTAATCTCATTAATTAATTAATGCTTTTAAATTAATGTAAGTGTATCA 997
DB 5264 AATTTAATAATCATCATATACATA---TATATTTAATAAATCAATAAATAATTTTAA 5208
QY 998 TTAAGATGAGAGAAACCAATATGCTGCTGCTGATTTTGAATTAATGTTTCTATGTT 1057
DB 5207 TACAACATACATAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5148
QY 1058 ACTTTTCTCAAGCTATATAAATACTTTGTAATGTAATTTGATGCTGGAATAATG 1117
DB 5147 AATATCCAAATAATATATTTTAAACAAAAAATAAATAAATAAATAAATAAATAA 5088
QY 1118 TGTATGATTCATTAATGAATTAATGATTTCAAGTCCAAATGCA--TCAATGAAT 1175
DB 5087 TAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 5028
QY 1176 TTAGTACAAACGTAACTCAAAATATTTCTTTTAAATTTTACACAAATTAATAA 1235
DB 5027 TATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4968
QY 1236 TATCTCTTATTTTAAATTTTACATAATTAATTAATTAATTAATTAATTAATTAAT 1295
DB 4967 TAAAAAATCAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4908
QY 1296 ACCAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1355
DB 4907 TCGATTTATTTTAACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4868
QY 1356 CTGATTTATTTTATTTTGTGCAATTTTCTTAATTTTGAAGTTAACCTTATATC 1415
DB 4847 ATCATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4789
QY 1416 TTGCTCAAACTAGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1475
DB 4788 CAATTCACAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4729
QY 1476 T 1476
DB 4728 T 4728

Search completed: November 25, 2003, 16:56:43
Job time : 1246 secs
```

Wed Nov 26 09:11:06 2003

us-09-831-083-1\_1.mpb

Page 10

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 13:13:48 ; Search time 96 Seconds  
(Without alignments)  
8197.772 Million cell updates/sec

Title: US-09-831-083-1  
Perfect score: 1783  
Sequence: 1 acccaactctgactcttga.....tagagcgaacgaactgaacc 1783

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: Issued\_Patents\_NA:\*  
2: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/Backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	73.8	4.1	20674	4	US-09-641-638-651
C 2	70	3.9	19124	2	US-08-487-8268-13
C 3	68.2	3.6	20674	4	US-09-641-638-651
C 4	67.6	3.8	665	2	US-08-883-795A-36
C 5	67	3.8	3095	6	5231168-1
C 6	66.6	3.7	665	2	US-08-883-795A-36
C 7	66.2	3.7	6243	2	US-09-056-075-1
C 8	62.6	3.5	19124	2	US-08-487-8268-13
C 9	61.6	3.5	668	3	US-08-998-416-535
C 10	61.6	3.5	6124	4	US-08-213-4198-3
C 11	61	3.4	168575	4	US-09-426-290-1
C 12	60.8	3.4	666	3	US-08-998-416-1117
C 13	60	3.4	53332	4	US-09-801-861-3
C 14	59	3.3	2614	4	US-09-004-056-1
C 15	58.8	3.3	827	4	US-08-998-416-535
C 16	58.8	3.3	58407	4	US-08-916-421B-2
C 17	57.8	3.2	10660	4	US-09-417-485D-5
C 18	57.8	3.2	53332	4	US-09-801-861-3
C 19	57.2	3.2	3701	3	US-08-845-258-10
C 20	57.2	3.2	3701	4	US-08-990-571-10
C 21	57.2	3.2	3701	4	US-08-723-142A-10
C 22	57.2	3.2	3701	4	US-09-528-783A-10
C 23	57.2	3.2	3701	4	US-09-569-098A-10
C 24	57	3.2	615	3	US-08-998-416-186
C 25	57	3.2	731	1	US-08-451-405A-2
C 26	57	3.2	2334	1	US-08-062-632-4
C 27	56.8	3.2	615	3	US-08-998-416-186

28	56.8	3.2	636	3	US-08-998-416-1137	Sequence 1137, App
29	56.8	3.2	837	3	US-08-998-416-288	Sequence 288, App
30	56.6	3.2	834	3	US-08-998-416-305	Sequence 305, App
C 31	56.6	3.2	837	3	US-08-998-416-288	Sequence 288, App
C 32	56.6	3.2	854	3	US-08-998-416-334	Sequence 534, App
C 33	55.8	3.1	834	3	US-08-998-416-305	Sequence 305, App
C 34	55.4	3.1	827	3	US-08-998-416-535	Sequence 535, App
C 35	55.4	3.1	1431	3	US-09-316-083-2	Sequence 2, App11
C 36	55.4	3.1	1431	4	US-09-933-700-2	Sequence 2, App11
C 37	55.2	3.1	6265	4	US-09-129-112-3	Sequence 3, App11
C 38	54.8	3.1	860	3	US-08-998-416-287	Sequence 287, App
C 39	54.8	3.1	168575	4	US-09-426-290-1	Sequence 1, App11
40	54.6	3.1	1511	1	US-07-991-867B-8	Sequence 8, App11
41	54.6	3.1	1511	1	US-08-107-755A-8	Sequence 8, App11
42	54.6	3.1	1511	2	US-08-544-332-8	Sequence 8, App11
43	54.6	3.1	1511	4	US-09-370-861A-8	Sequence 8, App11
C 44	54.2	3.0	731	1	US-08-451-405A-2	Sequence 2, App11
C 45	54.2	3.0	4818	3	US-08-817-926-27	Sequence 27, App11

## ALIGNMENTS

RESULT 1  
US-09-641-638-651/c  
; Sequence 651, Application US/09641638  
; Patent No. 6432648  
; GENERAL INFORMATION:  
; APPLICANT: Blumenfeld, Marta  
; APPLICANT: Bougueleret, Lydie  
; APPLICANT: Chumakov, Ilya  
; APPLICANT: Cohen, Annick  
; TITLE OF INVENTION: BALESTIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING  
; FILE REFERENCE: GENSET, 051CPI  
; CURRENT APPLICATION NUMBER: US/09/641, 638  
; PRIOR APPLICATION NUMBER: 2000-08-16  
; PRIOR FILING DATE: 1999-05-07  
; PRIOR APPLICATION NUMBER: US 09/275, 267  
; PRIOR FILING DATE: 1999-03-23  
; PRIOR APPLICATION NUMBER: US 60/119, 917  
; PRIOR FILING DATE: 1999-02-12  
; NUMBER OF SEQ ID NOS: 1304  
; SOFTWARE: Patent.pm  
; SEQ ID NO 651  
; LENGTH: 20674  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1123..3123  
; OTHER INFORMATION: 5' regulatory region  
; NAME/KEY: exon  
; LOCATION: 3124..3297  
; OTHER INFORMATION: exon 1  
; NAME/KEY: exon  
; LOCATION: 3871..4072  
; OTHER INFORMATION: exon 2  
; NAME/KEY: exon  
; LOCATION: 5552..5633  
; OTHER INFORMATION: exon 3  
; NAME/KEY: exon  
; LOCATION: 5758..5880  
; OTHER INFORMATION: exon 4  
; NAME/KEY: exon  
; LOCATION: 5996..6099  
; OTHER INFORMATION: exon 5  
; NAME/KEY: exon  
; LOCATION: 6349..6509  
; OTHER INFORMATION: exon 6



NAME/KEY: exon  
LOCATION: 7379..7522  
OTHER INFORMATION: exon 7  
NAME/KEY: exon  
LOCATION: 8645..8854  
OTHER INFORMATION: exon 8  
NAME/KEY: exon  
LOCATION: 12254..12340  
OTHER INFORMATION: exon 9  
NAME/KEY: exon  
LOCATION: 12854..13023  
OTHER INFORMATION: exon 10  
NAME/KEY: exon  
LOCATION: 13308..13429  
OTHER INFORMATION: exon 11  
NAME/KEY: exon  
LOCATION: 16567..16667  
OTHER INFORMATION: exon 12  
NAME/KEY: exon  
LOCATION: 16775..16945  
OTHER INFORMATION: exon 13  
NAME/KEY: exon  
LOCATION: 17063..17554  
OTHER INFORMATION: exon 14  
NAME/KEY: misc\_feature  
LOCATION: 17555..20674  
OTHER INFORMATION: 3'regulatory region  
NAME/KEY: allele  
LOCATION: 1128  
OTHER INFORMATION: 10-508-191 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 1182  
OTHER INFORMATION: 10-508-245 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 1559  
OTHER INFORMATION: 10-509-284 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 1579  
OTHER INFORMATION: 10-509-295 : deletion of C  
NAME/KEY: allele  
LOCATION: 1827  
OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTTT  
NAME/KEY: allele  
LOCATION: 2048  
OTHER INFORMATION: 10-511-62 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 2323  
OTHER INFORMATION: 10-511-337 : insertion of T  
NAME/KEY: allele  
LOCATION: 2341  
OTHER INFORMATION: 10-512-36 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 2623  
OTHER INFORMATION: 10-512-318 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 2832  
OTHER INFORMATION: 10-513-250 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 2844  
OTHER INFORMATION: 10-513-262 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 2934  
OTHER INFORMATION: 10-513-352 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 2947  
OTHER INFORMATION: 10-513-365 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 3802  
OTHER INFORMATION: 12-206-81 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 4062  
OTHER INFORMATION: 10-343-231 : deletion of C  
NAME/KEY: allele

LOCATION: 4088  
OTHER INFORMATION: 12-206-366 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 4109  
OTHER INFORMATION: 10-343-278 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 4170  
OTHER INFORMATION: 10-343-339 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 5903  
OTHER INFORMATION: 10-346-23 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6019  
OTHER INFORMATION: 10-346-141 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6141  
OTHER INFORMATION: 10-346-263 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 6183  
OTHER INFORMATION: 10-346-305 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 6338  
OTHER INFORMATION: 10-347-74 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6375  
OTHER INFORMATION: 10-347-111 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 6429  
OTHER INFORMATION: 10-347-165 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 6467  
OTHER INFORMATION: 10-347-203 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6484  
OTHER INFORMATION: 10-347-220 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6534  
OTHER INFORMATION: 10-347-271 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 6611  
OTHER INFORMATION: 10-347-348 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 7668  
OTHER INFORMATION: 10-348-391 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 8608  
OTHER INFORMATION: 10-349-47 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 8658  
OTHER INFORMATION: 10-349-97 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 8703  
OTHER INFORMATION: 10-349-142 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 8777  
OTHER INFORMATION: 10-349-216 : deletion of CTG  
NAME/KEY: allele  
LOCATION: 8785  
OTHER INFORMATION: 10-349-224 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 8926  
OTHER INFORMATION: 10-349-368 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12171  
OTHER INFORMATION: 10-350-72 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12429  
OTHER INFORMATION: 10-350-332 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 13341  
OTHER INFORMATION: 10-507-170 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 13492

OTHER INFORMATION: 10-507-321 : polymorphic base A or C  
 NAME/KEY: allele  
 LOCATION: 13524  
 OTHER INFORMATION: 10-507-353 : polymorphic base C or T  
 NAME/KEY: allele  
 LOCATION: 13535

Query Match 4.1%; Score 73.8; DB 4; Length 20674;  
 Best Local Similarity 51.3%; Pred. No. 8.8e-06;  
 Matches 254; Conservative 0; Mismatches 227; Indels 14; Gaps 3;

QY 1084 TTGTGATGCTAAATGATGCTGGAAGAAAATGTGAAGATCAATGAAATATG 1143  
 DB 11558 TTTTAAATATTAATATTTTCCCTAGCTATTAATATTAATATTAATATTAAT 11499  
 QY 1144 TATTCAAGTCCAAAATCCATCATAGAAATTTAGTACAAAACGTACTCAAAAATATT 1203  
 DB 11498 TAAATATTAATATTAATATTTTAAATTAATTAATTAATTAATTAATTAATTA 11439  
 QY 1204 CTCTATTTAAATTTTACACAAATATTA--AATATCTCTATTTTAAATTTTACAT 1261  
 DB 11438 AAATTTAATTTAATTTGAAGCAATTAATAATTAATTAATTAATTAATTAATTA 11379  
 QY 1262 AATATAATTTATGACCTGTCACCTTTAGAAATACCAACAATATTAATCTTATAT 1321  
 DB 11378 ATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11319  
 QY 1322 TTATCTTAATTAATTTGAAGATCTCTCAATATCTGATTTTATTTTATTTGTC 1381  
 DB 11318 TAAATTAAGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11259  
 QY 1382 TATTTCTTATGTTTGAAGTAACTTATATCTTGCAACTGATTAATCAATATAT 1441  
 DB 11258 TTTAATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 11199  
 QY 1442 GAGTTTGTGAAGACACATTTGACATCTTGAACATGCTTTAACCCT-----GTG 1493  
 DB 11198 AAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTA 11139  
 QY 1494 GAATGTAAAGTAAATTAATCAATTCAGAAATTAAGA-----CAATCTTAATTAATCTCT 1549  
 DB 11138 AAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTT 11079  
 QY 1550 TTGTCTTTTAAAAA 1564  
 DB 11078 TTAGCTATTAATAA 11064

## RESULT 2

US-08-487-826B-13/c  
 Sequence 13, Application US/08487826B  
 Patent No. 5993827

GENERAL INFORMATION:  
 APPLICANT: Sim, Kim L.  
 APPLICANT: Chitnis, Chetan  
 APPLICANT: Miller, Louis H.  
 APPLICANT: Peterson, David S.  
 APPLICANT: Su, Xin-zhaun  
 APPLICANT: Wellens, Thomas E.  
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
 NUMBER OF SEQUENCES: 45  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Knobbe Martens Olson & Bear  
 STREET: 620 Newport Center Drive 16th Floor  
 CITY: Newport Beach  
 STATE: California  
 COUNTRY: US  
 ZIP: 92660

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/487,826B  
 FILING DATE: 10-SEP-1993  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Israelson, Ned  
 REGISTRATION NUMBER: 29,655  
 REFERENCE/DOCKET NUMBER: NIH121.001CP1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 235-8550  
 TELEFAX: (619) 235-0176  
 INFORMATION FOR SEQ ID NO: 13:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 19124 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHEICAL: NO  
 ANTI-SENSE: NO  
 US-08-487-826B-13

Query Match 3.9%; Score 70; DB 2; Length 19124;  
 Best Local Similarity 44.3%; Pred. No. 4.7e-05;  
 Matches 286; Conservative 0; Mismatches 360; Indels 0; Gaps 0;

QY 770 TTTCGAGTTCATAGAGCACTTACAGACACTTTTACTAATAATCTACAAAGAGAG 829  
 DB 4861 TTCTCTTAATTTATTAATATATATATATATATATATATATATATATATATAT 4802  
 QY 830 ATTTTAACAATTAAGAGATATGAGATTAAGAGCAACATTTAAGGGAGTGA 889  
 DB 4801 AATAATTAATCAATATTTTATATCAATGACACTATGACTAATATATATATATTA 4742  
 QY 890 AATTAATGTTGTTGAACACACACTACCTTTAGTAATATATATATATATATATAT 949  
 DB 4741 ATACATACATATCTAATAAACAATATATATATATATATATATATATATATAT 4682  
 QY 950 CACATTAATTAATTTGCTTATTTTAAATTAATTAATTAATTAATTAATTAATTA 1009  
 DB 4681 ACATATATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4622  
 QY 1010 AAACCAATTAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1069  
 DB 4621 ATACTATATATATTTTATTAATGATATTTTTCATTTTAAAGTTTATTTTGTGAT 4562  
 QY 1070 GCTTATTAATTAATCTTGAATGCTAATTTGATGCTGGAAGAAAATGCTAATGAATTC 1129  
 DB 4561 CAATATTAATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4502  
 QY 1130 AATAGAAATTAATGATTTTCAAGTCCAAATCCATCATGAATTTAGTACAAAAGT 1189  
 DB 4501 TTAATTAATCTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4442  
 QY 1190 AACTCAAAATATCTCTTATTTTAAATTTTACAAATATTAATTAATTAATTTCTCTATTTT 1249  
 DB 4441 AACAATTAATTAATCTCATATCAATTAATTTTGAACAAATTAATTAATTAATTAAT 4382  
 QY 1250 AATTTTCAATTAATTAATTTTACACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1309  
 DB 4381 AAAAAAATTAATTAATGACATTTATGCTTATTTTATTAATTAATTAATTTTCT 4322  
 QY 1310 TACTAGATTTTATTTATCTTATATATTTTGAATCTGCAATATATGATATTTATTTT 1369  
 DB 4321 TTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4282  
 QY 1370 ATATTTGTGATATTTTCTTATGTTTGAATTAATTAATTAATTAATTAATTAATTA 1415  
 DB 4261 ATATATCTTACAAACATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4216

## RESULT 3

```
US-09-641-638-651
; Sequence 651, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bouguetere, Lydie
; APPLICANT: Chumakov, Ilya
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 651
; LENGTH: 20674
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1123..3123
; OTHER INFORMATION: 5'regulatory region
; NAME/KEY: exon
; LOCATION: 3124..3297
; OTHER INFORMATION: exon 1
; NAME/KEY: exon
; LOCATION: 3871..4072
; OTHER INFORMATION: exon 2
; NAME/KEY: exon
; LOCATION: 5552..5633
; OTHER INFORMATION: exon 3
; NAME/KEY: exon
; LOCATION: 5758..5880
; OTHER INFORMATION: exon 4
; NAME/KEY: exon
; LOCATION: 5996..6099
; OTHER INFORMATION: exon 5
; NAME/KEY: exon
; LOCATION: 6349..6509
; OTHER INFORMATION: exon 6
; NAME/KEY: exon
; LOCATION: 7379..7522
; OTHER INFORMATION: exon 7
; NAME/KEY: exon
; LOCATION: 8645..8854
; OTHER INFORMATION: exon 8
; NAME/KEY: exon
; LOCATION: 12254..12340
; OTHER INFORMATION: exon 9
; NAME/KEY: exon
; LOCATION: 12854..13023
; OTHER INFORMATION: exon 10
; NAME/KEY: exon
; LOCATION: 13308..13429
; OTHER INFORMATION: exon 11
; NAME/KEY: exon
; LOCATION: 16567..16667
; OTHER INFORMATION: exon 12
; NAME/KEY: exon
; LOCATION: 16775..16945
; OTHER INFORMATION: exon 13
; NAME/KEY: exon
; LOCATION: 17063..17554
; OTHER INFORMATION: exon 14

; NAME/KEY: misc_feature
; LOCATION: 17555..20674
; OTHER INFORMATION: 3'regulatory region
; NAME/KEY: allele
; LOCATION: 1128
; OTHER INFORMATION: 10-508-191 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1182
; OTHER INFORMATION: 10-508-245 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1559
; OTHER INFORMATION: 10-509-284 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 1570
; OTHER INFORMATION: 10-509-295 : deletion of C
; NAME/KEY: allele
; LOCATION: 1827
; OTHER INFORMATION: 10-510-173 : variable motif ATTGA or TTTTTT
; NAME/KEY: allele
; LOCATION: 2048
; OTHER INFORMATION: 10-511-62 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 2323
; OTHER INFORMATION: 10-511-337 : insertion of T
; NAME/KEY: allele
; LOCATION: 2341
; OTHER INFORMATION: 10-512-36 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 2623
; OTHER INFORMATION: 10-512-318 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2832
; OTHER INFORMATION: 10-513-250 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2844
; OTHER INFORMATION: 10-513-262 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 2934
; OTHER INFORMATION: 10-513-352 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 2947
; OTHER INFORMATION: 10-513-365 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 3802
; OTHER INFORMATION: 12-206-81 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 4062
; OTHER INFORMATION: 10-343-231 : deletion of C
; NAME/KEY: allele
; LOCATION: 4088
; OTHER INFORMATION: 12-206-366 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 4109
; OTHER INFORMATION: 10-343-278 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 4170
; OTHER INFORMATION: 10-343-339 : polymorphic base G or T
; NAME/KEY: allele
; LOCATION: 5303
; OTHER INFORMATION: 10-346-23 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6019
; OTHER INFORMATION: 10-346-141 : polymorphic base A or G
; NAME/KEY: allele
; LOCATION: 6141
; OTHER INFORMATION: 10-346-263 : polymorphic base G or C
; NAME/KEY: allele
; LOCATION: 6183
; OTHER INFORMATION: 10-346-305 : polymorphic base C or T
; NAME/KEY: allele
; LOCATION: 6338
; OTHER INFORMATION: 10-347-74 : polymorphic base A or G
; NAME/KEY: allele
```

LOCATION: 6375 : polymorphic base G or C  
OTHER INFORMATION: 10-347-111 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 6429 : polymorphic base C or T  
OTHER INFORMATION: 10-347-165 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 6467 : polymorphic base A or G  
OTHER INFORMATION: 10-347-203 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6484 : polymorphic base A or G  
OTHER INFORMATION: 10-347-220 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6534 : polymorphic base A or T  
OTHER INFORMATION: 10-347-271 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 6611 : polymorphic base A or G  
OTHER INFORMATION: 10-347-348 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 7668 : polymorphic base A or G  
OTHER INFORMATION: 10-348-391 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 8608 : polymorphic base C or T  
OTHER INFORMATION: 10-349-47 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 8658 : polymorphic base A or G  
OTHER INFORMATION: 10-349-97 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 8703 : polymorphic base G or C  
OTHER INFORMATION: 10-349-142 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 8777 : polymorphic base G or T  
OTHER INFORMATION: 10-349-216 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 8785 : polymorphic base C or T  
OTHER INFORMATION: 10-349-224 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 8926 : polymorphic base C or T  
OTHER INFORMATION: 10-349-368 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12171 : polymorphic base C or T  
OTHER INFORMATION: 10-350-72 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12429 : polymorphic base C or T  
OTHER INFORMATION: 10-350-332 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 13341 : polymorphic base A or G  
OTHER INFORMATION: 10-507-170 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 13492 : polymorphic base A or C  
OTHER INFORMATION: 10-507-321 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 13524 : polymorphic base C or T  
OTHER INFORMATION: 10-507-353 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 13535 : polymorphic base C or T

Query Match 3.8%; Score 68.2; DB 4; Length 20674;  
Best Local Similarity 46.5%; Pred. No. 0.0001;  
Matches 257; Conservative 0; Mismatches 293; Indels 3; Gaps 1;

781 CATAGAGCACTTACAGACCTTTACTTAATACTACAAAGAGAAATTTTACAC 840  
10976 CAATACTCTTGAACCTGCTGATGTACTATGTGAAAGACAAACAGCTTAAGATA 11035  
841 TTAGAGAGTATGAGAGTTAAAGACACACATTAAGGAGAGTGTAAATTAATG 900  
11036 AAAGAAACGATGTAAGACACAGCCCTTTTAAATAGCTAAGAAATTTATAT 11095  
901 TTGTACACACACTACTCTTATAGTAATATTAAGAAATTTGATCACTATTAAT 960  
11096 TTTTAATTTAATAATTTTATTTAATTTAAATTTTAATTTTAATTAATTAAT 11155  
961 TATTGCTTATTTAAATATATGATAAGTTGTA--TCATTAAGATGAGAAACCA 1017

11156 AAAATTTAATTTTAAATTAATAATTAATAATTAATAATTAATAATTAAT 11215  
1018 TAGTCCTGCTGTGAATTTTGAATATGTTTTCATGTTACTTTCTCAAGCCATAT 1077  
11216 TTAATTTTAAATTAACATTTTAATTTTATTTATTTAATTAATTAATTTAAT 11275  
1078 AAAAAGCTTGATGCTAATATGATGCTGAGAAAAATGCTAATGATTCATAGAAA 1137  
11276 TAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAA 11335  
1138 TTAATGATTTTCAAGTCCAAAATCCATCAATAGAAATTTAGTACAAAACGATCA 1197  
11336 TTAATTTAATTTAATTTTAAATTTAATTTAATTTAATTTAATTTAATTTAAT 11395  
1198 AATATCTCTTATTTAATTTTAAATTTTACACAAATTAATAATTTCTATTTAAATTT 1257  
11396 AATTTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 11455  
1258 CATATATATATTTATACCTGCTGACCTTTGAAATACACACATATTTAATTTGA 1317  
11456 TATTAATTTAATTTTAAATTTTAAATTTTAAATTTAATTTAATTTAATTTAAT 11515  
1318 TATTTATCTCTA 1330  
11516 TATTTAATTTTAA 11528

US-08-831-795A-36/c  
Sequence 36, Application US/08883795A  
Patent No. 5985607  
GENERAL INFORMATION:  
APPLICANT: Delcove, Genevieve  
APPLICANT: Awang, Gregor  
TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERSKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/883,795A  
FILING DATE: 27-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Michelle  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 7841-062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 361-1398  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 665 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: Rh 32  
US-08-883-795A-36

Query Match 3.8%; Score 67.6; DB 2; Length 665;  
Best Local Similarity 45.8%; Pred. No. 9.4e-05;  
Matches 271; Conservative 0; Mismatches 319; Indels 2; Gaps 1;

QY 1045 TGTGTTTATGTAAGTCTTTCTTCAAGCTTATTAATAAAGCTTGTATGCTAATGTATG 1104  
DB 596 TTTTATATTTTTCCTGCTGCTTTTAACTATTTATTAATAATGAAATTAATAAATGTA 537  
QY 1105 CTGGAATAAATGTAATGATTAATCAATAGAAATTTGTATTTCAAGTCGCAATTCGA 1164  
DB 536 ATATATATATCTTTATATTAATAAATGTAATTAATAATCTTAATTAATAAATGTA 477  
QY 1165 TCAATAGAAATTTAGTCAAAAGCTAAGCTCAAAATNTCTCTTATTTTAAATTTACA 1224  
DB 476 ATATATATCTTTATTAATAATGTAATTAATAAATGTAATTAATAAATGTAATTA 417  
QY 1225 CAATATTAATAATCTCTTATTTTAAATTTTCAATTAATAATTAATCACTGTCACC 1284  
DB 416 TAAATATGTAATTAATAAATGTAATTAATAAATGTAATTAATAAATGTAATTA 357  
QY 1285 TTTAGAAATACCAACCAATTAATTAATCTTAATTTTATTTCTTAATAATTTGAGATC 1344  
DB 356 TAAATATGTAATTAATAAATGTAATTAATAAATGTAATTAATAAATGTAATTA 297  
QY 1345 TCTCAATATATCTGAATTTTATTTTATTTGTCATTTTCTTATGTTTGAAGTGA 1404  
DB 296 TAAATATGTAATTAATAAATGTAATTAATAAATGTAATTAATAAATGTAATTA 237  
QY 1405 ACCCTTATATCTGTCGCAAACTAGTAATTAATTAATGAGTTTGTGAAGACATTTGAC 1464  
DB 236 TAAATATGTAATTAATAAATGTAATTTTAAATTTTAAATTTTAAATTTTAAAT 179  
QY 1465 ATCTTGAACATTTGTTTAACTCTGTCGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1524  
DB 178 TATTAATATGTAATTAATAAATGTAATTAATAAATGTAATTAATAAATGTAATTA 119  
QY 1525 ATGACCAATCTATTAATTAATCTGTTGCTTTTAAATAAAGTGCATGAATAATGCTCA 1584  
DB 118 TATTAATATGTAATTAATAAATGTAATTAATAAATGTAATTAATAAATGTAATTA 59  
QY 1585 TGTGTAAGCTAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1636  
DB 58 TATTAATATGTAATTAATAAATGTAATTAATAAATGTAATTAATAAATGTAATTA 7

RESULT 5  
5231168-1  
PATENT NO. 5231168  
APPLICANT: DIEGIEL, MORTEN; BORE, MARTIN; JENSEN, SOREN;  
INVENTOR: JENSEN, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.  
TITLE OF INVENTION: MALARIA ANTIGEN  
NUMBER OF SEQUENCES: 19  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/409,658  
FILING DATE: 18-SEP-1989  
SEQ ID NO: 1  
LENGTH: 3095  
5231168-1

Query Match 3.8%; Score 67; DB 6; Length 3095;  
Best Local Similarity 44.1%; Pred. No. 0.00014;  
Matches 280; Conservative 0; Mismatches 355; Indels 0; Gaps 0;  
QY 924 AAGTATTAATAGAAATGTAATCAATTAATTAATTTGCTTATTTAAATATG 983  
DB 2377 AAAAAAAAAAAAAAAAAATTAATAAATTTTCTTATTAATGTAATTAATTAAT 2436  
QY 984 ATAAAGTTGATCAATTAAGATTTGAGAAACCAATTAAGTCCGCTGCTGATTTTGAATTA 1043  
DB 2437 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2496  
QY 1044 TTGTTTCTATGTAATTTCTTCAAGCTATTAATAAATGTAATTAATGTAATTTAT 1103

DB 2497 ATTCAATGTAATTAATTAATTAATTAATTAATTAATTAATTTCTTTGCTGAT 2556  
QY 1104 GCTGGAATAAATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1163  
DB 2557 TAACTATCTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2616  
QY 1164 ATCAATGAATTTAGTCAAAAGCTAAGCTCAAAATNTCTCTTATTTTAAATTTACA 1223  
DB 2617 GAAAGTATGCAATTAATAAATAAATTTTCTGCTTATTAATTAATAAATTAATTA 2676  
QY 1224 ACAATATTAATAATCTCTTATTTTAAATTTTCAATTAATAATTAATTAATTAATTA 1283  
DB 2677 ATTTATGTAATTAATTAATAAATTAATTAATTAATTAATTAATTAATTAATTA 2736  
QY 1284 CTTTGAATTAACCAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2796  
DB 2737 AAAAAATTTTATTAATTAATTTTCTCAATAATTAATTAATTAATTAATTAATTA 2796  
QY 1344 CTTCAATATATCTGAATTTTATTTTATTTGTCATTTTCTTATGTTTGAAGTGA 1403  
DB 2797 CTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 2856  
QY 1404 AACCTTATATCTGTCGCAAACTAGTAATTAATTAATGAGTTTGTGAAGACATTTGA 1463  
DB 2857 AATAAATAATGAACAATAAATCAATGTCGATTTATTAATTTTGAATTAATAATTA 2916  
QY 1464 CATCTGAACATTTGTTTAACTCTGTCGTAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1523  
DB 2917 TATTAATTTTATTAATTTTAAATTTTAAATAATTAATTAATTAATTAATTAATTA 2976  
QY 1524 TATGACCAATCAATTAATTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1558  
DB 2977 AAAAAATGTAATTTATTAATGTCATTAATAATTTT 3011

RESULT 6  
US-08-883-795A-36  
Sequence 36, Application US/08883795A  
Patent No. 5985607  
GENERAL INFORMATION:  
APPLICANT: Delcove, Genevieve  
INVENTOR: Awang, Gregor  
TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BERESKIN & PARR  
STREET: 40 King Street West  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5H 3Y2  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/883,795A  
FILING DATE: 27-JUN-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Gravelle, Micheline  
REGISTRATION NUMBER: 40,261  
REFERENCE/DOCKET NUMBER: 7841-062  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 364-7311  
TELEFAX: (416) 364-1398  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 655 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: Rh 32  
US-08-883-795A-36

Query Match 3.7%; Score 66.6; DB 2; Length 665;  
Best Local Similarity 48.0%; Pred. No. 0.00015;  
Matches 282; Conservative 0; Mismatches 299; Indels 6; Gaps 3;

QY 959 ATTATGTCCTATTAAATATGATPAAGCTGTATCATTAAGATTGAGAAACCAAT 1018  
DB 14 AATGCTGTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 73  
QY 1019 AGTCCTGCTTGATTTTGAATTAATGTTTCTATGTAAGTCTTCTCAAGCTATATA 1078  
DB 74 ATTAATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 133  
QY 1079 A-AACTTGTAATGCTAAATGATGCGAAAAAATGTAATGAATTCATAGAA 1136  
DB 134 ATTAATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 193  
QY 1137 ATTATGTAATTCAGAACGCAAAATCCATAGAAATTTAGTACAAACGTAAGTCAA 1196  
DB 194 ATTAAATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 253  
QY 1197 AAAATATCTCTATTTTAAATTTTACAACATATAAAATATCTCTATTTTAAATTT 1256  
DB 254 ATTAATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 310  
QY 1257 ACATATATTAATTTACACCTGTCACCTTTAGAAATCCCAACATTTAATTAAGT 1316  
DB 311 ATTAATATCAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 369  
QY 1317 ATATTTATTTCTTAATTAATTTTGAATCTCTCAATATATGATTTTATTAATTTG 1376  
DB 370 TATAATATCAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTT 429  
QY 1377 TGTATATTTTCTTAATTTTAAAGTTAACCTTTAATCTGCTCAACAGTAATTC 1436  
DB 430 TATAATATCAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTT 489  
QY 1437 TATATGAGTTTGAAGGACACATGACATCTGAAACATGTTTAACTGTTGGA 1496  
DB 490 ATTAAAGTATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 549  
QY 1497 TGTAAAGGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1543  
DB 550 ATTCATATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 596

## RESULT 7

US-09-056-075-1  
Sequence 1, Application US/09056075

GENERAL INFORMATION:  
PATENT NO. 5955368  
APPLICANT: Johnson, Eric A.  
APPLICANT: Bradshaw, Marite  
TITLE OF INVENTION: Expression System for Clostridium  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Plinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/056,075  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.95238  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 608-251-5000  
TELEFAX: 608-251-9166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 6243 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: 3770..4013  
OTHER INFORMATION: /note="RP4 origin of DNA transfer (orit) from  
OTHER INFORMATION: plasmid RP4"  
US-09-056-075-1

Query Match 3.7%; Score 66.2; DB 2; Length 6243;  
Best Local Similarity 47.1%; Pred. No. 0.00022;  
Matches 303; Conservative 0; Mismatches 333; Indels 7; Gaps 3;

QY 917 CTTTGAAGTATTAATTAAGAAATGTAATGATCATCATTAATTAATTTGCTTAATTA 976  
DB 2768 CTTAACTCTTAATTAATTAATTAAGTAAGTCCCAATGGAACCTTAATTTTNG 2827  
QY 977 AATTGATTAAGTGTATCATTAATTAATTAAGTAAGTAAGTAAGTAAGTAAGTA 1036  
DB 2828 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2887  
QY 1037 TGAATTAATGTTTCTATGTTACTTTTCTCAAGCCATATAAACTTGTAAATGCTAA 1096  
DB 2888 TTTACTATTTCAATCCAAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2947  
QY 1097 ATGTATGCTG-GAAAAAATGTAATGTAATTAATTAATTAATTAATTAATTA 1154  
DB 2948 GTAAATATTTAATGATTTTACTTAATTTTGTATTAATTAATTAATTAATTAATTT 3007  
QY 1155 CCAAAATCCATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 1214  
DB 3008 TATAAATAAATCAAGAGCTTAATTAATTAATTAATTAATTAATTAATTAATTT 3067  
QY 1215 AATTTTACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTATC 1274  
DB 3068 AATTTTCAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 3127  
QY 1275 ACGCTGACCTTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 1334  
DB 3128 TCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 3185  
QY 1335 TTTGAGATCTCTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 1394  
DB 3186 ATTAAGTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 3245  
QY 1395 TTTGAGTTAACCTTTATTTTGTCAAACTAGTAATTAATTAATTAATTAATTTGTAAG 1454  
DB 3246 TTCCGCTTTA--TTAACTTAATTTTAAAGAAATTTTAACTTTTCAATTTTAAAG 3302  
QY 1455 ACATATGACATCTTGAAGATTTGTTTAACTTTGTAATTTTAAAGTAATTAATTAAC 1514  
DB 3303 ATTGATTAATTAATTTTAAATTAATTTTAACTTTTAACTTTTAAATTTTAAATTA 3362

QY 1515 ATTCAGATTATGACCATCTATTATATACCTCTTGTCTT 1557  
DB 3363 TTTTATTTATTTATTTATTTATTTATTTATTTTAAAGTTT 3405

RESULT 8  
US-08-487-826B-13

/ Sequence 13, Application US/08487826B  
/ Patent No. 5993827  
/ GENERAL INFORMATION:  
/ APPLICANT: Sim, Kim L.  
/ APPLICANT: Chitnis, Chetan  
/ APPLICANT: Miller, Louis H.  
/ APPLICANT: Peterson, David S.  
/ APPLICANT: Su, Xin-zhaun  
/ APPLICANT: Wellens, Thomas E.  
/ TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
/ TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS  
/ NUMBER OF SEQUENCES: 45  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: Knudde Warrens Olson & Bear  
/ STREET: 620 Newport Center Drive 16th Floor  
/ CITY: Newport Beach  
/ STATE: California  
/ COUNTRY: US  
/ ZIP: 92660  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: Patent in Release #1.0, Version #1.25  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/487,826B  
/ FILING DATE: 10-SEP-1993  
/ CLASSIFICATION: 435  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Israelson, Ned  
/ REGISTRATION NUMBER: 29,655  
/ REFERENCE/DOCKET NUMBER: NIH121.001CPI  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: (619) 235-8550  
/ TELEFAX: (619) 235-0176  
/ INFORMATION FOR SEQ ID NO: 13:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 19124 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: cDNA  
/ HYPOTHEICAL: NO  
/ ANTI-SENSE: NO  
/ US-08-487-826B-13

Query Match 3.5%; Score 62.6; DB 2; Length 19124;  
Best Local Similarity 47.4%; Pred. No. 0.0012;  
Matches 226; Conservative 0; Mismatches 244; Indels 7; Gaps 1;  
QY 906 ACCACCACTACCTTAGTATGATTAAGAAAATGTAACATACATTATATATTG 965  
DB 6547 ACTACTTTATATATATACACATACAGAAATATGATCTATATCAATTATATATA 6606  
QY 966 TCCTTATTTAAATTAAGATTAAGTGTATCATTAAGATGAAGAAACCAATAGCTC 1025  
DB 6607 TGAATATATATAATGATAGATATATAGATAGAGAGAAACGAAGCAATTTGCTCT 6666  
QY 1026 GTCTTGATTTGATTAATGTTTCTATGTTACTT-----TCTTCAAGCCTATATA 1078  
DB 6667 TTGTATCTCTAT 6726  
QY 1079 AAAACTTGTATGCTAAATTTGATGCTGCAAAAAATGTGTATGATTCATAGAAAT 1138  
DB 6727 CATATATATATGTTAATATATATATATATATATATATATATATATATATATAT 6786

QY 1139 TATGATTTCCAAAGTCCAAATCCATCATATAGAAATTTAGTACAAACGTAACGAAA 1198  
DB 6787 TTGATTTTGGATATTTTCTTCTCATTTATATATTTACTATATATATATATATAT 6846  
QY 1199 ATATTTCTTATTTTAAATTTTACACATATATATATTTCTTATTTTAAATTTTAC 1258  
DB 6847 AAAT 6906  
QY 1259 AAT 1318  
DB 6907 GATTAATATTTTGTGTTAGATTTTAAATTTTATATATATATATATATATATAT 6966  
QY 1319 ATTTATTTCTTAAATTTTGAATCTCTCAATATATCTATATTTATTTATTTT 1375  
DB 6967 ATATTTTAT 7023

RESULT 9  
US-08-998-416-595

/ Sequence 595, Application US/08998416  
/ Patent No. 6239264  
/ GENERAL INFORMATION:  
/ APPLICANT: Philippsen, Peter  
/ APPLICANT: Pohlmann, Rainer  
/ APPLICANT: Steiner, Sabine  
/ APPLICANT: Mohr, Christine  
/ APPLICANT: Wendland, Jurgen  
/ APPLICANT: Knechtle, Philipp  
/ APPLICANT: Reibschung, Corinne  
/ TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII  
/ TITLE OF INVENTION: AND USES THEREOF  
/ NUMBER OF SEQUENCES: 1152  
/ CORRESPONDENCE ADDRESS:  
/ ADDRESSEE: No. 6239264artis Corporation  
/ STREET: 3054 Cornwallis Road  
/ CITY: Research Triangle Park  
/ STATE: No 6239264th Carolina  
/ COUNTRY: USA  
/ ZIP: 27709  
/ COMPUTER READABLE FORM:  
/ MEDIUM TYPE: Floppy disk  
/ COMPUTER: IBM PC compatible  
/ OPERATING SYSTEM: PC-DOS/MS-DOS  
/ SOFTWARE: Patent in Release #1.0, Version #1.30  
/ CURRENT APPLICATION DATA:  
/ APPLICATION NUMBER: US/08/998,416  
/ FILING DATE: 24-DEC-1997  
/ CLASSIFICATION: 435  
/ PRIOR APPLICATION DATA:  
/ APPLICATION NUMBER: CH 0016/97  
/ FILING DATE: 31-DEC-1996  
/ ATTORNEY/AGENT INFORMATION:  
/ NAME: Weig, J. Timothy  
/ REGISTRATION NUMBER: 38,241  
/ REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
/ TELECOMMUNICATION INFORMATION:  
/ TELEPHONE: 919-541-8587  
/ TELEFAX: 919-541-8689  
/ INFORMATION FOR SEQ ID NO: 595:  
/ SEQUENCE CHARACTERISTICS:  
/ LENGTH: 658 base pairs  
/ TYPE: nucleic acid  
/ STRANDEDNESS: single  
/ TOPOLOGY: linear  
/ MOLECULE TYPE: DNA (genomic)  
/ ORIGINAL SOURCE:  
/ ORGANISM: PAG1408RP.  
/ US-08-998-416-595

Query Match 3.5%; Score 61.6; DB 3; Length 658;  
Best Local Similarity 47.7%; Pred. No. 0.0013;  
Matches 244; Conservative 0; Mismatches 264; Indels 4; Gaps 2;



QY 904 TAACCACACCTACCTTTAGATGATATAGAAATTTGTATCATCATATTAATAT 963  
DB 135 TAAGCACCACATATGAAATATGAAATTTTATGATCTTTGATATATACCATTTAACTT 194  
QY 964 TGGCTTATTTAAATATGATTAAGTGTATCTTTAAGATGAGAAACCAATATGCTC 1023  
DB 195 TATCTTGAATTAATTTTATTTATTTGTTATTTATTTT--ATTATTTATTTTATTTT 251  
QY 1024 TCGCTGATTTTGAATTTATTTGTTCTGATCTTTCTTCAAGCTATATTTAAAC 1083  
DB 252 ATATTTTATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTA 311  
QY 1084 TTTGTAATGCTAAATTTGATGCTGAGAAAAAATGCTAATGAAATTCATATGAAATATG 1143  
DB 312 TTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 371  
QY 1144 TATTTCAAGTCGCAAAATGCAATGCAATGAAATTTAG-TACAAAACGTAATCAAAATAT 1202  
DB 372 TTTATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 431  
QY 1203 TCTCTTATTTTAAATTTTACACAAATATTAATAATTTCTCTTATTTTAAATTTTACAATA 1262  
DB 432 CTATGAT 491  
QY 1263 ATATATTTTACACCTGTCACCTTTAGATATACCAACCAATATTTATTTATTTATTT 1322  
DB 492 ATATATATTTATTTTATTTGTAACAAATATATATATATATTTCTATTTATTTATTTAT 551  
QY 1323 TATTTCTATATATTTTATTTGATCTCTCAATATATCTGATATTTATTTATTTATTTGTCAT 1382  
DB 552 TTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 611  
QY 1383 ATTTTCTTATGTTTATGAGTTAAACCTTATAT 1414  
DB 612 TTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 643

## RESULT 10

US-06-213-419B-3/c  
; Sequence 3, Application US/08213419B  
; Patent No. 6333406  
; GENERAL INFORMATION:  
; APPLICANT: Inselburg, J. et al.  
; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM  
; TITLE OF INVENTION: AND USES THEREFOR  
; FILE REFERENCE: J11-002CNC  
; CURRENT APPLICATION NUMBER: US/08/213,419B  
; CURRENT FILING DATE: 1994-03-14  
; PRIOR APPLICATION NUMBER: US 07/870,506  
; PRIOR FILING DATE: 1992-04-17  
; NUMBER OF SEQ. ID NOS: 20  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ. ID NO 3  
; LENGTH: 6124  
; TYPE: DNA  
; ORGANISM: Plasmodium falciparum  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (2407)..(2439)  
; NAME/KEY: CDS  
; LOCATION: (2598)..(3404)  
; NAME/KEY: CDS  
; LOCATION: (3580)..(3720)  
; NAME/KEY: CDS  
; LOCATION: (3850)..(5835)  
; US-06-213-419B-3

Query Match 3.5%; Score 61.6; DB 4; Length 6124;  
Best Local Similarity 47.7%; Pred. No. 0.0017;  
Matches 210; Conservative 0; Mismatches 229; Indels 1; Gaps 1;  
QY 958 AATTATTTGCTTATTTAAAT 1017

DB 2579 AATGTTTGTATGCGTAAATAATATATATATATATATATATATATATATATATATATAT 2520  
QY 1018 TAGTCTGCTCTGATTTTGAATTTTGAATTTTCTATGTTACTTTTCTTCAAGCTATAT 1077  
DB 2519 TATTTGTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2460  
QY 1078 AAAAATTTGTAATGCTAAATTTGATGCTGAAAAAATGCTAATGAAATCAATAGAAA 1137  
DB 2459 TTTTCTTTTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2401  
QY 1138 TTAGTGATTTCAAAAGTCCAAATTCATATAGAAATTTAGTACAAAACGTAATCGAA 1197  
DB 2400 ATATGAAAT 2341  
QY 1198 AATATCTCTTATTTTAAATTTTACAAATATTAATAATTTCTCTTATTTTAAATTTTA 1257  
DB 2340 TATTTAATTTTCTTTTAT 2281  
QY 1258 CAAT 1317  
DB 2280 TTTTATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2221  
QY 1318 TATTTTATTTCTTAAATTTTGAATTTCTCTCAATATATCTGATATTTATTTATTTGCT 1377  
DB 2220 TAT 2161  
QY 1378 GTCATATTTTCTTATGTTT 1397  
DB 2160 TTTATTTATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 2141

## RESULT 11

US-09-426-290-1/c  
; Sequence 1, Application US/09426290  
; Patent No. 6410712  
; GENERAL INFORMATION:  
; APPLICANT: Berglind Ran Olafsdottir  
; APPLICANT: Jeffrey Guicher  
; TITLE OF INVENTION: HUMAN NARCOLEPSY GENE  
; FILE REFERENCE: 2345.2001-000  
; CURRENT APPLICATION NUMBER: US/09/426,290  
; CURRENT FILING DATE: 1999-10-25  
; NUMBER OF SEQ. ID NOS: 24  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ. ID NO 1  
; LENGTH: 168575  
; TYPE: DNA  
; ORGANISM: Homo Sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (21181)..(21403)  
; NAME/KEY: CDS  
; LOCATION: (95252)..(95430)  
; NAME/KEY: CDS  
; LOCATION: (101753)..(101996)  
; NAME/KEY: CDS  
; LOCATION: (110324)..(110439)  
; NAME/KEY: CDS  
; LOCATION: (124058)..(124278)  
; NAME/KEY: CDS  
; LOCATION: (127009)..(127130)  
; NAME/KEY: CDS  
; LOCATION: (128910)..(129139)  
; US-09-426-290-1

Query Match 3.4%; Score 61; DB 4; Length 168575;  
Best Local Similarity 49.1%; Pred. No. 0.0031;  
Matches 250; Conservative 0; Mismatches 250; Indels 9; Gaps 3;  
QY 923 TAAATATATATAGAAATTTGAATCATCATATATATATATATATATATATATATATATAT 982  
DB 109335 TAAAT 109276





TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAC1373RP  
US-08-998-416-535

Query Match 3.3%; Score 58.8; DB 3; Length 827;  
Best Local Similarity 49.1%; Pred. No. 0.0047;  
Matches 216; Conservative 0; Mismatches 217; Indels 7; Gaps 2;

QY 1040 ATTATGTTTCTATGTTACTTTCTTCAAGCCTATATATAAACTTTGATGCTAATT 1099  
DB 166 ATTTGCGCACTTATTTTATTTATTTAATTGATTAATTCATTAACATAAAACATT 225  
QY 1100 GATGCGTGGAAAAATGTAATGATTCATAGAAATATGATTTCAAGTCCAA 1159  
DB 226 TTAAATGTTAATAATTAATAAGAAATTTACTTATAGATAT-TTATTAAATAGTATT 284  
QY 1160 ATCCATCAATAGAAATTTAGTACAAAACGTACTCAAAATATTCCTTATTTAAATT 1219  
DB 285 ATTTATTTTATATTAATTAATACCATTTTATTATAAATAGATATTAAGTTTATTAA 344  
QY 1220 TACACCAATATTAATAATTTCTTATTTTAAATTTTACAATAATATTAATTCACCTG 1279  
DB 345 TATTAGTATATTAATTTTATTTATTTATTAATTTTATTTTCTTCAATGATATATAT 404  
QY 1280 TCACCTTTAGAAATACCAACCAATATTAATTAATTAATTTTATTTCTTAATATTTG 1339  
DB 405 AATTATTAATGATGCTTTCATTAATTTTATTTTATTTAGTCTGATTAATTCATTTAA 464  
QY 1340 AGATCTCCATATTTCTGATAT-T-TATTTATTTTGGTATATTTTCTTATG 1393  
DB 465 TAGCTACCTTTATTTGATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 524  
QY 1394 TTTTGAATTAACCTTATTAATTTGTCGAACTAGTAATTCATTAATTAATTAATTAAT 1453  
DB 525 AAGATATCTTAATCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATCTTT 584  
QY 1454 GACATTTGACATCTGAAA 1473  
DB 585 TTATATTTTAATCTTATTA 604

Search completed: November 25, 2003, 15:19:52  
Job time : 59 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: November 25, 2003, 13:13:48 / Search time 366 Seconds  
(without alignments)  
1150.541 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atccaactctgaccttga.....tagagcgatcaagctgaacc 1783

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2552756 segs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq.19Jun03:\*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1783	100.0	1909	21	AAA40223	V. faba sucrose bi
2	91.8	5.1	8056	25	ABJ10246	Haematopoietic cel
3	91.4	5.1	8056	25	ABJ10246	Haematopoietic cel
4	89	5.0	8056	25	ABJ10100	Haematopoietic cel
5	84.4	4.7	8056	25	ABJ10100	Haematopoietic cel
6	82.4	4.6	5930	24	ABL34492	Human immune syste
7	80	4.5	11996	24	ABL34492	Human immune syste
8	79.8	4.5	700	22	AAH93026	Human inflammatory

C	9	77.8	4.4	11996	24	ABL34493	Human metastasis a
	10	76.4	4.3	6641	24	ABL54336	Chemically treated
	11	76.4	4.3	6641	24	ABL32315	Human immune syste
	12	76.4	4.3	18154	24	ABL32254	Human immune syste
	13	76.2	4.3	19634	25	ABJ10015	Haematopoietic cel
	14	74.2	4.2	10329	24	ABL34123	Human immune syste
	15	74	4.2	7597	24	ABL33013	Human immune syste
	16	74	4.2	32392	24	ABL56203	Human immune syste
	17	73.8	4.1	20674	21	AAO58017	Arachidonic acid m
	18	73	4.1	19634	25	ABJ10061	Haematopoietic cel
	19	72.8	4.1	6189	22	AAJ46800	Tumour suppressor
	20	72.8	4.1	6189	24	ABL34593	Human metastasis a
	21	72.8	4.1	7667	22	AAJ46334	Tumour suppressor
	22	72.6	4.1	5689	22	AAJ45384	Chemically pretrea
	23	72.6	4.1	5689	22	AAJ46426	Tumour suppressor
	24	72.6	4.1	5689	24	ABK28226	DNA transcription
	25	72.6	4.1	14919	22	AAJ45506	Tumour suppressor
	26	72.2	4.0	6065	24	ABL32804	Human immune syste
	27	72	4.0	7025	24	ABK40059	Human chemically p
	28	71.8	4.0	7025	24	AAJ63350	Chemically pretrea
	29	71.8	4.0	5807	24	ABL33155	Human immune syste
	30	71.8	4.0	11996	24	ABL34493	Human immune syste
	31	71.4	4.0	5276	24	ABL32150	Human immune syste
	32	71.4	4.0	19087	24	ABL32793	Human immune syste
	33	71.4	4.0	19634	25	ABJ10162	Haematopoietic cel
	34	71	4.0	16258	24	ABL70376	Chemically treated
	35	71	4.0	16258	24	ABK40038	Human chemically p
	36	70.4	3.9	6255	24	ABL32960	Human immune syste
	37	70.4	3.9	8170	24	ABK28258	DNA transcription
	38	70.4	3.9	14820	24	ABN80146	Human chemically m
	39	70.2	3.9	8305	24	ABL33569	Human immune syste
	40	70	3.9	19124	18	AAT72882	Plasmodium var-7 p
	41	70	3.9	19124	21	AAZ98287	Plasmodium var-7 p
	42	69.8	3.9	3101	11	AAQ02047	Sequence encoding
	43	69.8	3.9	40862	24	ABL34073	Human immune syste
	44	69.8	3.9	50000	24	ABL56201	Human immune syste
	45	69.8	3.9	50000	24	ABL56202	ABPV genome fragm

#### ALIGNMENTS

RESULT 1	
ID AAA40223	standard; DNA, 1909 BP.
AC AAA40223;	
XX	
DT 02-NOV-2000	(first entry)
XX	
DE V. faba sucrose binding protein DNA fragment.	
XX	
KW Soybean; sucrose binding protein; SBP; plant seed; transgenic plant;	
KW seed-specific expression; ds.	
XX	
OS Vicia faba.	
XX	
FH	
FT Key	Location/Qualifiers
FT CDS	1784..1909
FT	/*tag= a
FT	/product= "SBP"
FT	/partial
PN	
XX	
MO200026386-A2.	
PD 11-MAY-2000.	
XX	
PF 27-OCT-1999;	99WO-DE03432.
XX	
PR 04-NOV-1998;	98DE-1052195.
XX	
PA (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.	
XX	

Human metastasis a  
Chemically treated  
Human immune syste  
Human immune syste  
Human immune syste  
Haematopoietic cel  
Human immune syste  
Human immune syste  
Arachidonic acid m  
Haematopoietic cel  
Tumour suppressor  
Human metastasis a  
Tumour suppressor  
Chemically pretrea  
Tumour suppressor  
DNA transcription  
Tumour suppressor  
Human immune syste  
Human chemically p  
Chemically pretrea  
Human immune syste  
Human immune syste  
Human immune syste  
Haematopoietic cel  
Chemically treated  
Human chemically p  
Human immune syste  
DNA transcription  
Human chemically m  
Human immune syste  
Plasmodium var-7 p  
Plasmodium var-7 p  
Sequence encoding  
Human immune syste  
ABPV genome fragm  
ABPV genome fragm

PI Heim U, Weber H;  
XX MPI; 2000-365631/31.  
DR P-PSDB; AAB10028.

XX Expression cassette for expressing genes in plant seeds, useful for  
PT producing enzymes or pharmaceuticals, includes the promoter from a  
PT sucrose-binding protein-related gene

XX Disclosure; Fig 1; 24pp; German.

XX This invention describes a novel expression cassette (I) for expressing  
CC genes in plant seeds which comprises (i) the promoter of the  
CC sucrose-binding protein (SBP)-like seed protein (II), (ii) optionally the  
CC sequence for a signal peptide, particularly from (II), (iii) the gene  
CC (III) to be expressed and (iv) a 3'-termination sequence. (I) and a  
CC plasmid containing the expression cassette (IV) are used for the  
CC expression of homologous or heterologous genes in the seeds of  
CC transformed plants, particularly genes that alter the storage properties  
CC and germination capacity of the seeds. Alternatively, transgenic plants  
CC that express altered or new products in their seeds are selected, grown  
CC to establish stable lines and the resulting products (e.g. enzymes,  
CC pharmaceuticals or proteins that contain essential amino acids) are  
CC extracted. Also insertion of (III) in the antisense orientation may be  
CC used to reduce or switch off expression of particular genes. (I) provide  
CC seed-specific expression (in cotyledons and endosperm), with stable  
CC expression at high level. They provide an overall increase in expression  
CC rate, improve utilization of the developmental period of the seed and  
CC can overcome the effect of co-suppression. This sequence encodes a  
CC fragment of the Vicia faba (soybean) sucrose binding protein (SBP) which  
CC is described in the method of the invention.

XX Sequence 1909 BP; 638 A; 317 C; 281 G; 673 T; 0 other;

Query Match 100.0%; Score 1783; DB 21; Length 1909;  
Best Local Similarity 100.0%; Pred. No 3,5e-296;  
Matches 1783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATCCAACTTCGTGATCTTGAATCTCTGTTCCAAACATGTTGGAAGAGCTTAAAGACT 60
DB 1 ATCCAACTTCGTGATCTTGAATCTCTGTTCCAAACATGTTGGAAGAGCTTAAAGACT 60
QY 61 TTTGAGAAAGCTTTGATCAGCTTTGAGACTTTCTTGAATTAAGCTTGGAAAGCTGA 120
DB 61 TTTGAGAAAGCTTTGATCAGCTTTGAGACTTTCTTGAATTAAGCTTGGAAAGCTGA 120
QY 121 TTTGAGAAAGCTTTGATCAGCTTTGAGACTTTCTTGAATTAAGCTTGGAAAGCTGA 180
DB 121 TTTGAGAAAGCTTTGATCAGCTTTGAGACTTTCTTGAATTAAGCTTGGAAAGCTGA 180
QY 181 AAATTTATGAGTACCTTCAAGTTTCAAGAGTCTTCAAGATTCGATTTGAAGTTC 240
DB 181 AAATTTATGAGTACCTTCAAGTTTCAAGAGTCTTCAAGATTCGATTTGAAGTTC 240
QY 241 CATATTTTAAAGAGAGTTCTGTTCCGAATGCTTGAAGTCTTGAAGTTCGAACT 300
DB 241 CATATTTTAAAGAGAGTTCTGTTCCGAATGCTTGAAGTCTTGAAGTTCGAACT 300
QY 301 CTGAGTTCAGAGTCTTCCAGATCAATTCGATCAATGCTTGAAGTTCGAACT 360
DB 301 CTGAGTTCAGAGTCTTCCAGATCAATTCGATCAATGCTTGAAGTTCGAACT 360
QY 361 CTGAGTTCAGAGTCTTCCAGATCAATTCGATCAATGCTTGAAGTTCGAACT 420
DB 361 CTGAGTTCAGAGTCTTCCAGATCAATTCGATCAATGCTTGAAGTTCGAACT 420
QY 421 TTTCAAGCAAACTTGAATTTGACTTTCTTAATTAACAACTTCAATTAATCAATGT 480
DB 421 TTTCAAGCAAACTTGAATTTGACTTTCTTAATTAACAACTTCAATTAATCAATGT 480
QY 481 CTGATGAATGATGATTTGGAATTTGATGATGATGATGATGATGATGATGATGATG 540
DB 481 CTGATGAATGATGATTTGGAATTTGATGATGATGATGATGATGATGATGATGATG 540

```

```

QY 541 AGTGGCAATGACCAATTTGCTCTTGCCAAATTCGAACTTAATTAATGATGATGATG 600
DB 541 AGTGGCAATGACCAATTTGCTCTTGCCAAATTCGAACTTAATTAATGATGATGATG 600
QY 601 CTGCAAACTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
DB 601 CTGCAAACTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660
QY 661 TTTTGTGATCAACACAAAGAAATCTGTTTTCATGCGGACATGACATTAACATTA 720
DB 661 TTTTGTGATCAACACAAAGAAATCTGTTTTCATGCGGACATGACATTAACATTA 720
QY 721 AACACCACTTCAATGCAAGAGTGAATTAAGAGAAATGTCAGTTCCTTTCGAGT 780
DB 721 AACACCACTTCAATGCAAGAGTGAATTAAGAGAAATGTCAGTTCCTTTCGAGT 780
QY 781 CATAAGCAACCTTACAGACCTTTTACTTAATTAATCAAAAGAGAGATTTTAACAC 840
DB 781 CATAAGCAACCTTACAGACCTTTTACTTAATTAATCAAAAGAGAGATTTTAACAC 840
QY 841 TTAGAGAAATGAGAGTGAATTAAGAGCAACATTAAGGGGAGTGTAAATTAATGTG 900
DB 841 TTAGAGAAATGAGAGTGAATTAAGAGCAACATTAAGGGGAGTGTAAATTAATGTG 900
QY 901 TTTGATCAACCACTTACCTTTAGTATTAATTAAGAAATTTGATCAATCAATTAAT 960
DB 901 TTTGATCAACCACTTACCTTTAGTATTAATTAAGAAATTTGATCAATCAATTAAT 960
QY 961 TATGTCCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020
DB 961 TATGTCCTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1020
QY 1021 TCTGCTGATGATTTTGAATTTATTTTCTTAATTAATTTCTTCAAGCTTAAATA 1080
DB 1021 TCTGCTGATGATTTTGAATTTATTTTCTTAATTAATTTCTTCAAGCTTAAATA 1080
QY 1081 AACTTGAATGCTTAATTTGATGCTGAAATTAATGATTAATTAATTAATTAATTA 1140
DB 1081 AACTTGAATGCTTAATTTGATGCTGAAATTAATGATTAATTAATTAATTAATTA 1140
QY 1141 TGGATTTTCAAGTCAAAATTCATCAATTAAGAAATTTGATCAATTAATTAATTA 1200
DB 1141 TGGATTTTCAAGTCAAAATTCATCAATTAAGAAATTTGATCAATTAATTAATTA 1200
QY 1201 ATTCTCTTATTTAAATTTTCAACAAATTAATTAATTAATTAATTAATTAATTA 1260
DB 1201 ATTCTCTTATTTAAATTTTCAACAAATTAATTAATTAATTAATTAATTAATTA 1260
QY 1261 TAAATTAATTTATCACTGCTCACTTTGAATTAATCAACAAATTAATTAATTAAT 1320
DB 1261 TAAATTAATTTATCACTGCTCACTTTGAATTAATCAACAAATTAATTAATTAAT 1320
QY 1321 TTTATTTCTTAATTAATTTGAGATCTTCAATTAATTAATTAATTAATTAATTA 1380
DB 1321 TTTATTTCTTAATTAATTTGAGATCTTCAATTAATTAATTAATTAATTAATTA 1380
QY 1381 TTTATTTCTTAATTAATTTGAGATCTTCAATTAATTAATTAATTAATTAATTA 1440
DB 1381 TTTATTTCTTAATTAATTTGAGATCTTCAATTAATTAATTAATTAATTAATTA 1440
QY 1441 TGAATTTGAGAGCAACATTTGACATTTGAAACATTTGTTTAACTTTGAGAAAT 1500
DB 1441 TGAATTTGAGAGCAACATTTGACATTTGAAACATTTGTTTAACTTTGAGAAAT 1500
QY 1501 AAAGGTAATTAACAACTTGAATTAATTAATTAATTAATTAATTAATTAATTA 1560
DB 1501 AAAGGTAATTAACAACTTGAATTAATTAATTAATTAATTAATTAATTAATTA 1560
QY 1561 AAAAGTGTGATGAATTAATGATGATGATGATGATGATGATGATGATGATGATG 1620
DB 1561 AAAAGTGTGATGAATTAATGATGATGATGATGATGATGATGATGATGATGATG 1620

```

QY 1621 AATTCATTTCCAGATGCTAGAACTGCGACTACGAATATTAGTCATTAAGACACGTATG 1680  
 DB 1621 AATTCATTTCCAGATGCTAGAACTGCGACTACGAATATTAGTCATTAAGACACGTATG 1680  
 QY 1681 TTAACACACGTCCTCCCTGATGTTTTCGCAATATTCGCTCTCTTTCTTTCTTCAC 1740  
 DB 1681 TTAACACACGTCCTCCCTGATGTTTTCGCAATATTCGCTCTCTTTCTTTCTTCAC 1740  
 QY 1741 GTATTAACCAATGAATATTATTAAGACGATCAAGCTGAACC 1783  
 DB 1741 GTATTAACCAATGAATATTATTAAGACGATCAAGCTGAACC 1783

RESULT 2  
 ID AB210246/c standard; DNA; 8056 BP.  
 XX AB210246;  
 AC AB210246;  
 DT 16-JAN-2003 (first entry)  
 XX  
 DE Haematopoietic cell proliferation disorder related DNA sequence #386.  
 XX  
 KM Human; haematopoietic cell proliferation disorder; cytostatic;  
 KM gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
 KM cytosine methylation state; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WC200277272-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 26-MAR-2002; 2002MO-EP03401.  
 XX  
 PR 26-MAR-2001; 2001US-278333P.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;  
 PI Olek A, Pipenbrock C, Adorjan P, Grabs G, Jesche R, Leu E;  
 PI Lewin A, Lipscher E, Mader S, Model F, Meller V, Otto T;  
 PI Pelet C, Schwobe I, Ziebarth H;  
 XX  
 DR WPI; 2003-018942/01.  
 XX  
 PT Detecting and differentiating between hematopoietic cell proliferative  
 PT disorders, comprises contacting a target nucleic acid with a reagent  
 PT that distinguishes between methylated and non-methylated CpG  
 PT dinucleotides -  
 XX  
 PS Claim 28; SEQ ID 386; 1179p; English.  
 XX  
 CC The present invention describes a method for detecting and  
 CC differentiating between haematopoietic cell proliferative disorders  
 CC associated with at least 1 gene and/or their regulatory regions in a  
 CC subject. The method comprises contacting a target nucleic acid in a  
 CC biological sample obtained from the subject with at least 1 reagent,  
 CC which distinguishes between methylated and non-methylated CpG  
 CC dinucleotides within the target nucleic acid. AB203861 to AB211118  
 CC represent specifically claimed nucleotide sequences from the present  
 CC invention. Oligonucleotides from the present invention can be used: for  
 CC differentiating between healthy haematopoietic cells and proliferative  
 CC disorder haematopoietic cells; for differentiating between acute  
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
 CC determining the cytosine methylation state and/or single nucleotide  
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
 CC related sequences and their complements; and as primers for the  
 CC amplification of haematopoietic cell proliferation disorder related  
 CC DNA sequences. The nucleotide sequences from the present invention can  
 CC also be used for detecting a predisposition to, differentiation between  
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of  
 CC haematopoietic cell proliferative disorders. The present method enables

CC a highly specific classification of haematopoietic cell proliferative  
 CC disorders allowing for improved and informed treatment of patients.  
 XX

SO Sequence 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;  
 Query Match 5.1%; Score 91.8; DB 25; Length 8056;  
 Best Local Similarity 45.0%; Pred. No. 6.7e-07;  
 Matches 481; Conservative 0; Mismatches 572; Indels 16; Gaps 3;

QY 707 ACGCATTAACATTAACACCACTTCATTGAGAGGATGAGAGCAAGCAATGCGAGTT 766  
 DB 2848 AATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2789  
 QY 767 ACCCTTCGAGTTCATTAAGACCACTTACAGACACTTTTACCTAAATCTACAAAGAG 826  
 DB 2788 AATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2729  
 QY 827 AAGATTTTACCACTTGAAGAGTAAAGGAGTTAAAGACACACATTACGGGAGTG 886  
 DB 2728 AATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2669  
 QY 887 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 946  
 DB 2668 TTTAAATTTTAAATTTTAACTTAC-----ATTTTTATTTTATTTTAAATTTAT 2614  
 QY 947 CATCACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1006  
 DB 2613 AATTAATTAATTTTTCATTTATTTTATTTTATTTTAAATTTTATTTAAATTTTAA 2554  
 QY 1007 AAGAAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTCT 1066  
 DB 2553 AATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2494  
 QY 1067 CAAGCCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1126  
 DB 2493 TATTTTAAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 2434  
 QY 1127 TTCAATTAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAA 1186  
 DB 2433 AATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2374  
 QY 1187 CGTAACCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAAAT 1246  
 DB 2373 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2314  
 QY 1247 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1306  
 DB 2313 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTAAATTTTAAAT 2254  
 QY 1307 TAAATCTAGATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1366  
 DB 2253 TAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTT 2194  
 QY 1367 TTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1426  
 DB 2193 TTAATTAATTTTAAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2134  
 QY 1427 AGTAATTTA-----ATTAATGAAGTTTGAAGACACATTGACATCTTGAACATT 1477  
 DB 2133 AAAAAATTAATTAATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTT 2074  
 QY 1478 GGTTTTAACTGTTGAGATGTTAAAGGATTAATTAATTAATTAATTAATTAATTAATTAAT 1537  
 DB 2073 TTTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 2014  
 QY 1538 AATTAATCTGCTGCTTTTAAATTAATTAATTAATTAATTAATTAATTTTAAATTTTAAAT 1597  
 DB 2013 TAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1954  
 QY 1598 TGTCTGTGGCGCGTGATATCAATTCATTCAGATGGTAAAGTGGCACTAGAGAA 1657  
 DB 1953 TTTTATTTTATTTATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAA 1894



QY 1658 TAATAGTCATAGACACGATGTTAACAAGCTCCCTGCATGTTTTCATATAT 1717  
 DB 1893 AAAAAAAAAAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 1834  
 QY 1718 TCGGCTCTTCCTT--TTCTCAGCTTAAACAAATGATATAT 1764  
 DB 1833 AACAAATTTTATTATTATTATTAAATTAATTAATTAATTAAT 1785

RESULT 3  
 AB210246  
 ID AB210246 standard; DNA; 8056 BP.  
 XX  
 AC AB210246;  
 DT 16-JUN-2003 (first entry)  
 XX  
 DE Haematopoietic cell proliferation disorder related DNA sequence #246.  
 XX  
 KM Human; haematopoietic cell proliferation disorder; cytostatic;  
 KM gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
 KM cytosine methylation state; gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO20027272-A2.  
 XX  
 PD 03-OCT-2002.  
 XX  
 PF 26-MAR-2002; 2002MO-EP03401.  
 XX  
 PR 26-MAR-2001; 2001US-278333P.  
 XX  
 PA (EPIS-) EPIGENOMICS AG.  
 XX  
 PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;  
 PI Olek A, Piepenbrock C, Ascorjan P, Grabs G, Lesche R, Leu F;  
 PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;  
 PI Peter C, Schwope I, Ziebarth H;  
 XX  
 DR WPI; 2003-018942/01.  
 XX  
 PT Detecting and differentiating between haematopoietic cell proliferative  
 PT disorders, comprises contacting a target nucleic acid with a reagent  
 PT that distinguishes between methylated and non-methylated CpG  
 PT dinucleotides -  
 XX  
 PS Claim 28; SEQ ID 386; 117bp; English.  
 XX  
 CC The present invention describes a method for detecting and  
 CC differentiating between haematopoietic cell proliferative disorders  
 CC associated with at least 1 gene and/or their regulatory regions in a  
 CC subject. The method comprises contacting a target nucleic acid in a  
 CC biological sample obtained from the subject with at least 1 reagent,  
 CC which distinguishes between methylated and non-methylated CpG  
 CC dinucleotides within the target nucleic acid. AB209861 to AB21118  
 CC represent specifically claimed nucleotide sequences from the present  
 CC invention. Oligonucleotides from the present invention can be used: for  
 CC differentiating between healthy haematopoietic cells and proliferative  
 CC disorder haematopoietic cells; for differentiating between acute  
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
 CC determining the cytosine methylation state and/or single nucleotide  
 CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
 CC related sequences and their complements; and as primers for the  
 CC amplification of haematopoietic cell proliferation disorder related  
 CC DNA sequences. The nucleotide sequences from the present invention can  
 CC also be used for detecting a predisposition to, differentiation between  
 CC haematopoietic cell proliferative disorders. The present method enables  
 CC a highly specific classification of haematopoietic cell proliferative  
 CC disorders allowing for improved and informed treatment of patients.  
 XX  
 SQ \*sequence, 8056 BP; 3711 A; 0 C; 371 G; 3974 T; 0 other;

Query Match 5.1%; Score 91.4; DB 25; Length 8056;  
 Best Local Similarity 45.7%; Pred. No. 7.9e-07;  
 Matches 355; Conservative 0; Mismatches 421; Indels 1; Gaps 1;

QY 799 ACACCTTTACTAAATACACAGAGAGATTTTACACACTTACAGAGATATGAGAG 858  
 DB 626 AAAAAATTTTAAAAATATTTTAAATTAATTTTAAAAAATTTTAAAAATTAATA 685  
 QY 859 TTAAGACGACACATTAAGGGAGTGTAAATTAATGATGTGTACACACCACTACCT 918  
 DB 686 ATATTAATAAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATA 745  
 QY 919 TTAGTATGATTAATTAAGAAATTTGTAATCATCATTAATTAATTTGCTTAATTAATA 978  
 DB 746 TTTTAAATTTTAAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATAAT 805  
 QY 979 TTAAG-ATAAGTGTATCATTAAGATTTGAATTTGAACCAATTAAGTCCGCTGATTTT 1037  
 DB 806 AAATTTAAATTAATGATTTTAAATTAATTTTAAATTAATTAATTAATTTTATTTT 865  
 QY 1038 GAATTAATTTGTTTCTAATGTTACTTTCTTCAAGCTTATTAATAACTTTGTAATGCTAA 1097  
 DB 866 TTTATGATTAATAAATTTTGTAAATTTTAAATTTTAAATTTTAAATTTTGTGTTTT 925  
 QY 1098 TTGATAGCTGGAATAAATGATGATGATTAATTAATTAATTAATTAATTAATTAATGCA 1157  
 DB 926 GTAAATTAATTTTAAATTAATTAATTTTATTTTAAATTAATTAATTAATTAATTAATA 985  
 QY 1158 AAATTCATCAATTAAGAAATTTAGTACAAAGCTTACCAAAATTTCTCTTAATTAAT 1217  
 DB 986 TTAATTAATTTTAAAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTGAAA 1045  
 QY 1218 TTACAAACAATTAATAAATTTCTTAAATTTTAAATTTTAAATTAATTAATTAATTAATCAAC 1277  
 DB 1046 TTTAATTAATGTAATAAATTAATAAATTTGTTTAATTTGTTTAAATTTTAAATTTTAAATA 1105  
 QY 1278 TGTCACTTTAAGATTAACCAACAATTAATTAATTAATTAATTAATTTTAAATTTTAAATTT 1337  
 DB 1106 TGTATTAATTAATTTGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTAT 1165  
 QY 1338 TGAGATCTCAATTAATGATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1397  
 DB 1166 TTTATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1225  
 QY 1398 AGAGTTAACCTTAATTAATTTGTTCAAACTTAATTAATTAATTAATTAATTTGTAAGAGACA 1457  
 DB 1226 TTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1285  
 QY 1458 CATTAACATCTTGAACATTTGTTTAACTTTGTAATTTGTAATTTGTAATTTGTAATTTGTAATTT 1517  
 DB 1286 AATATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1345  
 QY 1518 CAGATTAATGACCATTAATTAATTAATTTGCTTTGCTTTTAAATAAGTGTGACATGA 1574  
 DB 1346 ATTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTGA 1402

RESULT 4  
 AB210100/c  
 ID AB210100 standard; DNA; 8056 BP.  
 XX  
 AC AB210100;  
 DT 16-JUN-2003 (first entry).  
 XX  
 DE Haematopoietic cell proliferation disorder related DNA sequence #240.  
 XX  
 KM Human; haematopoietic cell proliferation disorder; cytostatic;  
 KM gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;  
 KM cytosine methylation state; gene; ds.  
 XX  
 OS Homo sapiens.

```

XX  WO200277272-A2.
PN  XX
XX  03-OCT-2002.
FD  XX
XX  26-MAR-2002; 2002WO-EP03401.
PF  XX
XX  26-MAR-2001; 2001US-278333P.
PR  XX
XX  (EPIC-) EPIDENOMICS AG.
PA  XX
XX  Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI  Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI  Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
PI  Pelet C, Schwöpe I, Ziebarth H;
XX  WPI; 2003-018942/01.
DR  XX
XX  Detecting and differentiating between hematopoietic cell proliferative
PT  disorders, comprises contacting a target nucleic acid with a reagent
PT  that distinguishes between methylated and non-methylated CpG
PT  dinucleotides -
XX  XX
PS  Claim 28; SEQ ID 240; 117bp; English.
XX  XX
CC  The present invention describes a method for detecting and
CC  differentiating between hematopoietic cell proliferative disorders
CC  associated with at least 1 gene and/or their regulatory regions in a
CC  subject. The method comprises contacting a target nucleic acid in a
CC  biological sample obtained from the subject with at least 1 reagent,
CC  which distinguishes between methylated and non-methylated CpG
CC  dinucleotides within the target nucleic acid. AB209861 to AB211118
CC  represent specifically claimed nucleotide sequences from the present
CC  invention. Oligonucleotides from the present invention can be used: for
CC  differentiating between healthy hematopoietic cells and proliferative
CC  disorder hematopoietic cells; for differentiating between acute
CC  lymphocytic leukaemia and acute myelogenous leukaemia; as probes for
CC  determining the cytosine methylation state and/or single nucleotide
CC  polymorphisms (SNPs) of hematopoietic cell proliferation disorder
CC  related sequences and their complements; and as primers for the
CC  amplification of hematopoietic cell proliferation disorder related
CC  DNA sequences. The nucleotide sequences from the present invention can
CC  also be used for detecting a predisposition to, differentiation between
CC  subclases, diagnosis, prognosis, treatment and/or monitoring of
CC  hematopoietic cell proliferative disorders. The present method enables
CC  a highly specific classification of hematopoietic cell proliferative
CC  disorders allowing for improved and informed treatment of patients.
XX  XX
SQ  Sequence 8056 BP; 3711 A; 371 C; 371 G; 3603 T; 0 other;
Query Match 5.0%; Score 89; DB 25; Length 8056;
Best Local Similarity 48.2%; Pred. No. 2e-06;
Matches 309; Conservative 0; Mismatches 330; Indels 2; Gaps 2;
QY 929 TTTAAGAAATGTGATCATCATTAATATGTCCTATTAAATATGATTA 988
DB 2543 TTTTATTAATTAATTAATTTTATTTTTCGTTAAATTTT-TAATATTTTAA 2485
QY 989 GTTGATTCATTAAGATTGAGAAACCAATAGTCCTGCTGATTGTAATTTGTT 1048
DB 2484 ACGAAATTAATTAATAATATATATAAATAAATAAATAATTTTAAATAATA 2425
QY 1049 TTCTATGCTACTTTTTCGACGCTATATAAAACCTTTGTAATGCTAAATGTAAGCTGG 1108
DB 2424 AATTATTTTATTTTATTTTATTTTGAATAATTAATTAATTAATTTTAAATA 2365
QY 1109 AAAAAATGCTAATGATTCATAGAAATTAATGATTTTCAAGCCAAATTCATCA 1168
DB 2364 AAAAAATTAATTAATAATTAATAATTAATTTTAAATCAAAATTAATTAATTTT 2305
QY 1169 TGAATTTTATGTAACAAACGTAACTCAAAATTTCTTATTTTAAATTTTACACA-A 1227
DB 2304 TAAAAATTAATTAATTAATAATTTTATTTTAAATTTTAAATTAATTAATAATA 2245

```

```

QY 1228 TATAAAAATATCTCTTATTTTAAATTTTACAAATATATATTTACACGTCACCTTT 1287
DB 2244 AATTTTATTAATAATTAATTTTATTTTATTTTTCGTTTATTTTAAATTT 2185
QY 1288 AGAATACCAACCAATATTAATTAATTTTATTTTCTTAATTAATTTGAGATCTT 1347
DB 2184 TTAATTAATAATTTTATTTTATTTTATTTTATTTTAAATAATAATAATAATTA 2125
QY 1348 CAATATATCTGATTTATTTATTTATTTGTCATATTTTCTTATGTTTAAAGTTAAC 1407
DB 2124 TAAATTAATTAATTAATTTTAAATTTATTTTAAATTTTAAATAATTTTAAATA 2065
QY 1408 CTTAATCTGCTGCAACTAGTAATTCATATATAGTTTGTGAAGACACATTCACATC 1467
DB 2064 ATTTTATTAATAATAATAATAATTTTAAATAATATTTTAAATATATTTTAAATTT 2005
QY 1468 TTGAACATTTGCTTTTAACTTTGGAATGTTAAAGTAATAAATTCGAAATTATG 1527
DB 2004 ATTAATTTTAAATAATAATAATTAATTAATTTTATTTTAAATAATTAATTTT 1945
QY 1528 ACCATCTAATTAATATACCTTCCTTCTTTTAAATAAGTGT 1568
DB 1944 TTTATATTAATAATTAATTTTATTTTATTTTAAATAATTTT 1904
RESULT 5
AB210100
ID AB210100 standard; DNA; 8056 BP.
XX
XX AB210100;
AC
XX
DT 16-JAN-2003 (first entry)
XX
DE Hematopoietic cell proliferation disorder related DNA sequence #240.
XX
KW Human; hematopoietic cell proliferation disorder; cytostatic;
KW gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
KW cytosine methylation state; gene; ds.
XX
OS Homo sapiens.
XX
PN WO200277272-A2.
XX
XX 03-OCT-2002.
PD
XX
PF 26-MAR-2002; 2002WO-EP03401.
XX
PR 26-MAR-2001; 2001US-278333P.
XX
PA (EPIC-) EPIDENOMICS AG.
XX
PI Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
PI Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Leu E;
PI Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
PI Pelet C, Schwöpe I, Ziebarth H;
XX
DR WPI; 2003-018942/01.
XX
XX Detecting and differentiating between hematopoietic cell proliferative
PT disorders, comprises contacting a target nucleic acid with a reagent
PT that distinguishes between methylated and non-methylated CpG
PT dinucleotides -
XX
PS Claim 28; SEQ ID 240; 117bp; English.
XX
XX The present invention describes a method for detecting and
XX differentiating between hematopoietic cell proliferative disorders
XX associated with at least 1 gene and/or their regulatory regions in a
XX subject. The method comprises contacting a target nucleic acid in a
XX biological sample obtained from the subject with at least 1 reagent,
XX which distinguishes between methylated and non-methylated CpG
XX dinucleotides within the target nucleic acid. AB209861 to AB211118

```

CC represent specifically claimed nucleotide sequences from the present  
 CC invention. Oligonucleotides from the present invention can be used: for  
 CC differentiating between healthy hematopoietic cells and proliferative  
 CC disorder hematopoietic cells; for differentiating between acute  
 CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
 CC determining the cytosine methylation state and/or single nucleotide  
 CC polymorphisms (SNPs) of hematopoietic cell proliferation disorder  
 CC related sequences and their complements; and as primers for the  
 CC amplification of hematopoietic cell proliferation disorder related  
 CC DNA sequences. The nucleotide sequences from the present invention can  
 CC also be used for detecting a predisposition to, differentiation between  
 CC subclasses, diagnosis, prognosis, treatment and/or monitoring of  
 CC hematopoietic cell proliferative disorders. The present method enables  
 CC a highly specific classification of hematopoietic cell proliferative  
 CC disorders allowing for improved and informed treatment of patients.

XX Sequence 8056 BP; 3711 A; 371 C; 3603 T; 0 other;

Query Match 4.7%; Score 84.4; DB 25; Length 8056;  
 Best Local Similarity 46.7%; Pred. No. 1.2e-05;  
 Matches 369; Conservative 0; Mismatches 416; Indels 5; Gaps 3;

QY 779 TTCTAAGACGCACTTACAGACACTTTACTAAATCTCAAGAGAGAGATTTACAC 838  
 Db 274 TTAATTAATAAATAATTTTATTTTAAATTTTAAATAATTAATAAATAA 333  
 QY 833 ACTTGAAGAGTAATGAGAGTT-AAAGGCAACATTAGGGGAGTGAATTAAT 897  
 Db 334 ATTAATTAATAAATAATTAATAAATAAATAAATAAATAAATAATTAATA 393  
 QY 898 GTGTGTACACACACCTCTTGTAGTATTAAGAAATGTATCATCATTAAT 957  
 Db 394 AAATTAATTAATAATTAATTTTATTTTATTTTAAATAAATAAATAAATA 453  
 QY 958 AATTATTCCTTATTTAAATTTATGATAAGTGTATCATTAAGATTGGAACCAA 1017  
 Db 454 AATTATTTTAAATAATTAATTAATAAATAAATAAATAAATAAATAATTTT 512  
 QY 1018 TAGTCTGCTCTGATTTTGAATTAATTTTGTATGATTAATTTTCTTAACCTATAT 1077  
 Db 513 TAAATTAATAATTAATTAATAATTAATAATTTTATTAATTAATAACCTTAT 572  
 QY 1078 AAAAATTTTATGCTTAATTTATGATGCTGAAAAAATGTTATGATTAATTCAT 1137  
 Db 573 AAAAATTTTATTTAATAATTAATAATTAATAAATAATTAATAATTTTAA 632  
 QY 1138 TTAGGTATTTCAAAAGTCCAAATTCATCATAGAAATTTAGTACAAAACGTAC 1197  
 Db 633 TTTAAATAATTTTAAATAATTAATAATTAATAAATAAATAAATAAATAA 692  
 QY 1198 AATATTCCTTATTTAATTTTACACAAATTAATAATTTCTTATTTTAAATTT 1257  
 Db 693 AAAAATTTTATTTAATAATTTAATAATTTTATTAATTAATAATTTTAA 752  
 QY 1258 CAATATATTAATTTATCAGCTGACCTTGAATTAACCAACAAATTAATTAAC 1317  
 Db 753 AATTAAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTA 812  
 QY 1318 TATTTATTTCTTAATTAATTTGAGATCTCTCAATATTCGATTTATTTATTT 1377  
 Db 813 TAAAGATTTATTAATAATTTTATTAATAAATAATGTTTTTTTATTTTACG 872  
 QY 1378 GTCATATTTCTTAATTTAGTTAGATTACCTTA--TATCTGTCGAACACTGATTC 1434  
 Db 873 ATTAATAATTTGCTAATTTAATAATTTAATAATTTTTCGTCGTTGCTATTA 932  
 QY 1435 AATAATAGAGTTTGGAAGGACATTCGATCTGAAACATTTGTTTACCTTGTTG 1494  
 Db 933 ATTTTTTTAAATAATTTTATTTTATTTTAAACGAATTAATAATTAATTAAT 992  
 QY 1495 AATGTTAAGGTAATTAATAACATTCAGATTAATGACATTAATAATTAATCTTGC 1554  
 Db 993 TTTTAAATAATTAATTTTAAATAATTTTATTTTATTTTATTTGGAATTAAT 1052

QY 1555 TTTTAAAAA 1564  
 Db 1053 ACGTAAAAATA 1062

RESULT 6  
 ID ABLJ2517/C  
 ABJ2517 standard; DNA; 5930 BP.

AC ABLJ2517;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 490.

XX Human; immune system disease; cytosine methylation; antileukemic;  
 KW antileukemic; antileukemic; antileukemic; antileukemic;  
 KW antileukemic; antileukemic; antileukemic; antileukemic;  
 KW antileukemic; antileukemic; antileukemic; antileukemic;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

PF 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-103259.

PK 01-SEP-2000; 2000DE-1043826.

PA (EPIC-) EPIGENOMICS AG.

PI Olek A. Piepenbrock C. Berlin K;

DR WPI: 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation

XX Claim 1; SEQ ID NO 490; 32bp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 5930 BP; 2122 A; 24 C; 912 G; 2872 T; 0 other;

Query Match 4.6%; Score 82.4; DB 24; Length 5930;  
 Best Local Similarity 48.3%; Pred. No. 2.7e-05;  
 Matches 230; Conservative 0; Mismatches 246; Indels 0; Gaps 0;

QY 936 AAAATTTAATTCATCATTAATTAATTTGCTTATTTAAATTAATGATAAGTTGAT 995  
 Db 5149 AAAAATTAATAAATAATTAATAAATAATTTACTTAAATTTATTAATAAATAATTT 5090  
 QY 996 CATTAGATTGAAGAAACCAATATAGCTCGTCTGATTTGAAATTAATTTGTTCTATG 1055  
 Db 5089 TAAAAAATTTACTCATCAATTAATTAATAAATCTCAAACTCCAACTTATTAATCTAT 5030  
 QY 1056 TTACTTTCTTCAAGCCTATTAATAAATCTTGAATGCTTAATTTGATGCTGAAAAA 1115

Db 5029 TTATATATTTTCACAAACATATTTAAATAACAAAAATACAAATTTAAATAAAAAAA 4970  
QY 1116 TGTGTAATGAAATTCATAGAAATTAATGTAATTTAAAGTCGCAATTCATCATAGAAAT 1175  
Db 4969 ACAATATATATTTACTTTATATTAATAATTTCTAATTAATAATAAAAACTTAAAAACCAT 4910  
QY 1176 TTAGTACAAAACGTAAGTCAAAAAATATCTCTTAATTTTAAATTTTACACATATTAATA 1235  
Db 4909 AACAAATATATATTAATATATTAATAAAAAATATCTTAAATAATACATTAATAATA 4850  
QY 1236 TATCTCTTAATTTAAATTTTACATATATATTTATTCACCTGCTCCTTTGAAATACC 1295  
Db 4849 AACTCTAAGTAAATAATTAATAATAATAATTAACCATATATATATTTTAAATATA 4790  
QY 1296 ACCAACAATATTAATCTTATAGATATTTTATCTTAATTAATTTGAGATCTGCATATAT 1355  
Db 4789 AAAATATATATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 4730  
QY 1356 CTGATATTTATTTATTTATTTGTCGATATTTCTTATGTTTGAAGTTAACCTTTA 1411  
Db 4729 ATTAACTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4674

## RESULT 7

ABL34492  
ID ABL34492 standard; DNA; 11996 BP.

AC ABL34492;

DT 26-MAR-2002 (first entry)

DE Human metastasis associated gene SRO ID NO: 45.

KW Metastasis associated gene; cytostatic; gene therapy; cancer;  
cytosine methylation; gene; ds.

OS Homo sapiens.

PN MO200177376-A2.

PD 18-OCT-2001.

PF 06-APR-2001; 2001WO-EP03970.

PR 06-APR-2000; 2000DE-1019058.

PR 07-APR-2000; 2000DE-1019173.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

PA (EPiG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-010922/01.

PT New nucleic acid derived from chemically treated metastasis genes,  
also for diagnosis of cancers by analysis of cytosine methylation,  
also for treatment -

PS Claim 1, SEQ ID NO 45; 23bp + Sequence Listing; English.

CC The present invention provides a number of human metastasis associated  
CC genes which are modified by cytosine methylation. The sequences can be  
CC used in the diagnosis and treatment of cancer. The present sequence is  
CC one of the genes of the invention.

SQ Sequence 11996 BP; 2961 A; 261 C; 3028 G; 5746 T; 0 other;

Query Match 4.5%; Score 80; DB 24; Length 11996;  
Best Local Similarity 49.0%; Pred. No. 7.1e-05;

Matches 242; Conservative 0; Mismatches 250; Indels 2; Gaps 1;

QY 912 ACTACTTAGTAGTATTAAGAAATTTGTAATCATCATCTTAATTAATTTGT-CCT 969

Db 929 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 988  
QY 970 TATTTAAATTTATGATAAAGTTGATTCATTTAGATTTGAGAAAAACCAATAGTCGCT 1029  
Db 989 TATATATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 1048  
QY 1030 TGAATTTTGAATTTATTTGTTTCTATGTTACTTTTCTCAAGCTATTTAAAACTTTGTA 1089  
Db 1049 TATATTTATTTATGATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAG 1108  
QY 1090 ATGCTAAATTTGATGCTGCAAAAAATGTCATGATTTCAATTTAGATTTAGTATTC 1149  
Db 1109 ATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 1168  
QY 1150 AAAGTCCAAATCCATCCATTTAGTAAATTTAGTAAATTTAGTAAATTTAGTAAATTT 1209  
Db 1169 ATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTA 1228  
QY 1210 TTTAAATTTTACCAAAATTTAAATAATTTCTATTTTAAATTTTCAATTAATTAAT 1269  
Db 1229 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATA 1288  
QY 1270 TTATGACCTGTCACCTTTAGAAATTCACCAAAATTTATTTATTTATTTATTTATCT 1329  
Db 1289 TTAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATA 1348  
QY 1330 AATATTTTGAAGATCTCCCAATTTATCGATTTATTTATTTATTTATTTATTTATTTCT 1389  
Db 1349 ATTATTTATTTGTTGGATTAATTTTGGGTTGTTTTTTTTTTTTTTTTTTTTTTGA 1408  
QY 1390 TATGTTTGAAGTT 1403  
Db 1409 GATGAGTTTGT 1422

## RESULT 8

AAH93026/C  
ID AAH93026 standard; DNA; 700 BP.

AC AAH93026;

DT 09-OCT-2001 (first entry)

DE Human inflammatory bowel disease related gene fragment ICR3119a.

KW Human; inflammatory bowel disease; Crohn's disease; ulcerative colitis;  
single nucleotide polymorphism; SNP; chromosome 19p13; paternity test;

KW chromosome 5q11-33; forensic test; gene therapy; ds.

OS Homo sapiens.

PN WO200142511-A2.

PD 14-JUN-2001.

PF 11-DEC-2000; 2000WO-US33632.

PR 10-DEC-1999; 99US-0170257.

PR 10-APR-2000; 2000US-0196046.

PA (WHEB) WHITEHEAD INST BIOMEDICAL RES.  
(ELLIT-) ELLIPSIS BIOTHERAPEUTICS CORP.

PI Daly M, Hudson JV, Lander ES, Rioux J, Siminovitch K;

DR WPI; 2001-367874/36.

PT Testing for the presence of polymorphisms associated with inflammatory  
bowel disease, using a hybridization assay -

PS Disclosure; Page 384; 463bp; English.

XX

CC The present invention describes a method for detecting the presence of  
 CC polymorphisms associated with inflammatory bowel diseases such as  
 CC ulcerative colitis and Crohn's disease. The methods can be used to detect  
 CC the presence of genetic polymorphisms associated with inflammatory bowel  
 CC disease and correlating their occurrence with disease states. They may be  
 CC used in this way for phenotypic correlations, forensics, paternity  
 CC testing, medicine and genetic analysis. The present sequence is a gene  
 CC containing a polymorphic site described in the exemplification of the  
 CC invention.

XX Sequence 700 BP; 370 A; 13 C; 14 G; 302 T; 1 other;

Query Match 4.5%; Score 79.8; DB 22; Length 700;

Best Local Similarity 48.5%; Pred. No. 6.6e-05;

Matches 315; Conservative 0; Mismatches 323; Indels 12; Gaps 3;

919 TTATGATTAAGTGTGATTCATTAAGTGAAGAAACCAATAGCTGCTGATTTTG 1038  
 637 TTAAT 578  
 1039 AATTATGCTTCTGATCTTCTTCTCAAGCTATATATAAAGCTTGAATGCTAAAT 1098  
 577 AAAAT 518  
 1099 TGTATGCTGAAAAAATGCTGAATGATTCATTAAGAAATATGATTTCAAGTCAA 1158  
 517 TAAAT 458  
 1159 AATCATCATTAAGAAATTTAGTACAAACGTAACCTCAAAATATCTTATTTAAAT 1218  
 457 TAT-----AAT 403  
 1219 TTACAACAATATATATATCTTATTTTAAATTTTAAATATATATATATAT 1278  
 402 TAT 343  
 1279 GTACACCTTGAATGCCCAACAATATATATATATATATATATATATATAT 1338  
 342 TAT 285  
 1339 GAGATCTCTCAATATATCTGATTTTATATATATATATATATATATATAT 1398  
 284 TAT 225  
 1399 GAGTAAACCTTATATCTGCTCAAACTAGTAATCAATATATGAGTTTGTGAAGAC 1458  
 224 TATTTA-----TATATATATATATATATATATATATATATATATATAT 170  
 1459 ATTGACATCTGAAACATATGTTTAACTTGTGAAAGTTAAAGTAATAAAGATTC 1518  
 169 AAT 110  
 1519 AGAATTAAGACCATATATATATATATATATATATATATATATATATATAT 1568  
 109 ATATTTAT 60

RESULT 9

ABL34493/c

ID ABL34493 standard; DNA; 11996 BP.

XX ABL34493;

XX 26-MAR-2002 (first entry)

XX Human metastasis associated gene SFQ ID NO: 46.

XX Metastasis associated gene; cytostatic; gene therapy; cancer;

KX "cytosine methylation; gene; ds.

XX Homo sapiens.

XX WC200177376-A2.

XX 18-OCT-2001.

XX 06-APR-2001; 2001WO-EP03970.

XX 06-APR-2000; 2000DE-1019058.

XX 07-APR-2000; 2000DE-1019173.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A. Piepenbrock C. Berlin K;

XX WPI; 2002-010922/01.

XX New nucleic acid derived from chemically treated metastasis genes,  
 PT useful for diagnosis of cancers by analysis of cytosine methylation,  
 PT also for treatment

XX Claim 1; SEQ ID NO 46; 23pp + Sequence Listing; English.

CC The present invention provides a number of human metastasis associated  
 CC genes which are modified by cytosine methylation. The sequences can be  
 CC used in the diagnosis and treatment of cancer. The present sequence is  
 CC one of the genes of the invention.

XX Sequence 11996 BP; 2966 A; 261 C; 3041 G; 5728 T; 0 other;

Query Match 4.4%; Score 77.8; DB 24; Length 11996;

Best Local Similarity 50.1%; Pred. No. 0.00017;

Matches 248; Conservative 0; Mismatches 242; Indels 5; Gaps 2;

912 ACTACCTTGTAGTATATATATAGAAATTTGTAATGATCATATATATATATGCT 969  
 11068 AT 11009  
 970 TATTTAAAT 1029  
 11008 TAT 10949  
 1030 TGAATTTGAAATATATGTTTCTATGTTTCTTCAAGCTATATATAAAGCTTGT 1089  
 10948 TAAAT 10889  
 1090 ATGCTAAAT 1149  
 10888 AT 10829  
 1150 AAAGTCCAAATGCAATGAAATTTTGTGCAAAACGTAACCTCAAAATATCTCTTA 1209  
 10828 ATATATTTTATATATATATATATATATATATATATATATATATATAT 10772  
 1210 TTTTAAATTTTAAACAATATATAAATATCTTATTTTAAATTTTAAATATATAT 1269  
 10771 AT 10712  
 1270 TTAATACCTGCACTTGAATATATATATATATATATATATATATATATAT 1329  
 10711 AT 10652  
 1330 AATATATTTGAGATCTCTCAATATATATATATATATATATATATATATAT 1389  
 10651 AT 10592  
 1390 TATGTTTATAGCTTA 1404  
 10591 TAAATATATATCTTA 10577

RESULT 10  
 ABL54336 standard; DNA; 6641 BP.  
 ID ABL54336  
 XX  
 AC ABL54336;  
 XX  
 DT 29-JUL-2002 (first entry)  
 XX  
 DE Chemically treated apoptosis gene complementary to gene #18.  
 XX  
 KW Apoptosis; HIV; Bloom syndrome; cardiopathy;  
 KM neurodegenerative disorder; Herpes simplex virus; renal ischaemia;  
 KM amyotrophic lateral sclerosis; cancer; ds.  
 OS  
 XX Unidentified.  
 XX  
 PN WO200177164-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PE 06-APR-2001; 2001WO-EP03969.  
 XX  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPig-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-017444/02.  
 XX  
 PT Chemically modified sequences of genes associated with apoptosis are  
 PT useful to determine methylation patterns of genomic DNA samples for  
 PT diagnosis of associated diseases such as cancer -  
 XX  
 PS Claim 1; Seq ID #36; 24bp; English.  
 XX  
 CC This invention relates to chemically pre-treated DNA of genes  
 CC associated with apoptosis. The nucleic acids are used to allocate  
 CC patients for specific therapy for HIV infection, Bloom syndrome,  
 CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus  
 CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours  
 CC and cancers. This nucleotide sequence represents a chemically  
 CC treated apoptosis gene. Even SEQ ID numbers are the complementary  
 CC DNA strands to the odd SEQ ID numbers. The sequence data for this  
 CC patent is not represented in the printed specification but is based on  
 CC information supplied by the European patent office.  
 XX  
 SQ Sequence 6641 BP; 2495 A; 22 C; 984 G; 3140 T; 0 other;  
 Query Match 4.3%; Score 76.4; DB 24; Length 6641;  
 Best Local Similarity 45.9%; Pred. No. 0.00028;  
 Matches 372; Conservative 0; Mismatches 431; Indels 7; Gaps 3;  
 QY 755 AAGTGGACGTAACCTCTTGTGAGCTGATGAGAGCACTTACAGACCTTTTAACTAAAT 814  
 Db 2561 AT 2620  
 QY 815 ACTACAAAGAGAGAGATTTTAAACACTGAGAGAGTAAGGAGTTAAAGACACACAT 874  
 Db 2621 ATATTTAAAGTAATTAATTTTATTTTAAATATAGAGATATATATTAAGTGTATAG 2680  
 QY 875 TAAGGGGAGGTGTTAAATTAATGCTGTAACACACACTTACCTTAGTAGATTTATA 934  
 Db 2681 TGAGAAAGTTAGATATTTTGTGTAATTTTAAAGATATATATTTAGTTTAT 2740  
 QY 935 GAAATTTGTAATATACATTAATTAATTTGCTTATTTAAATTAATATAGTAAAGTTGA 994  
 Db 2741 TTAAATATAGATAGTATTTTATATATATATTTTAAAGAAATATGTTAGATAT 2800

QY 995 TCATTAAGATTAGAGAAACCAATAGTCTGCTGTGATTTTGAATTAATGTTTCAT 1054  
 Db 2801 TGAATTAATTTGGTTTGAAGATATGATTTATTTTACGTTATTTGAGTTATAT 2860  
 QY 1055 GTTACTTTTCTCAAGCCATATAAACTTTGA--ATGCTAATTTGATGCTGGAA 1111  
 Db 2861 GTTAAATATGAATTTATTTATTTTGAAGATATATGAGATTAATTAATGATTA 2920  
 QY 1112 AAAATGTATGAATTCATTAATGAATTAATG--TATTTCAAGTCCAAATCCATCAAT 1169  
 Db 2921 GTTATATTTAAATTAATTAATTTTATGATTTTGTGATTTAAAAATGATTTAT 2980  
 QY 1170 AAAAAATTTAGTCAAAAGCTACTAAAAATTTCTTATTTTAAATTTTCAACATA 1225  
 Db 2981 ATTAAGTATGATTAATTTATTTATTAATGATTAATTAATTAATTAATTAATTA 3040  
 QY 1230 TAAAAATATCTCTTATTTTAAATTTTCAATATATATATTAACCTGACCTTTAG 1289  
 Db 3041 GTAATTTATTTATTTATATATGATTTTAAATTTAGTAATTTATTTATTAATTA 3100  
 QY 1290 AATACACCAACAAATTTATATCTTAGATTTTATCTTAATTAATTTGAGATCTCA 1349  
 Db 3101 ATTAGTAAATTTATTTATTTATTTATGATTAATTAATTTAGTAATTTATTTAT 3160  
 QY 1350 ATATATCGATATTTATTTTA--TATTTGTCATTTTCTTATGTTTGAAGTTAAC 1407  
 Db 3161 AATTAATTTAGTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3220  
 QY 1408 CTATATCTGTGTCAAAGTATGATTAATCAATATATGATTTTGAAGACATGACATC 1467  
 Db 3221 ATTAATTTTATTTTAAAGATTTTAAATTTAGTATTTAGATTTTATTAATTAATTA 3280  
 QY 1468 TTGAACATTTGTTTAACTTTGTGAGATGTTAAAGTATTAACATTTCAATTTATG 1527  
 Db 3281 GTAGATATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 3340  
 QY 1528 ACCATCTATTAATTAATCTCTTGTGCTTT 1557  
 Db 3341 TATATGATTTAGGTTTAAATTTATTTT 3370

RESULT 11  
 ABL32315  
 ID ABL32315 standard; DNA; 6641 BP.  
 XX  
 AC ABL32315;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 288.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KM antiarteriosclerotic; antianemic; cyrostatic; noctropical;  
 KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KM antidiabetic; anticancer; antidiabetic; antiparasitic;  
 KM antineoplastic; cancer; eye disease; arteriosclerosis; anaemia;  
 KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KM gene; ds.  
 OS  
 XX Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PE 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPig-) EPIGENOMICS AG.  
 XX

PI Olek A, Piepenbrock C, Berlin K;  
XX  
XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful  
PT for diagnosis and treatment of diseases associated with abnormal  
PT cytosine methylation -

XX PS Claim 1; SEQ ID NO 288; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated  
CC genes which are modified by the methylation of cytosines. The sequences  
CC can be used in the diagnosis and treatment of immune system disorders,  
CC including eye diseases such as retinopathy, neovascular glaucoma and  
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
CC diseases. The present sequence is a gene of the invention.

XX SQ Sequence 6641 BP; 2495 A; 22 C; 984 G; 3140 T; 0 other;

XX Query Match 4.3%; Score 76.4; DB 24; Length 6641;

XX Best Local Similarity 45.9%; Pred. No. 0.00028;

XX Matches 372; Conservative 0; Mismatches 431; Indels 7; Gaps 3;

QY 755 AATGTCACCTTACCTTCGCGCTTCATAGACACTTACGACACTTTACTAAAT 814  
DB 2561 AT 2620

QY 815 ACTACAG 874  
DB 2621 AT 2680

QY 875 TAAAGGAG 934  
DB 2681 TGAAG 2740

QY 935 GAAATGAT 994  
DB 2741 TTAAT 2800

QY 995 TCATTAAGATGAG 1054  
DB 2801 TGAATTAAT 2860

QY 1055 GTTACTTTCTTCAAGCCAT 1111  
DB 2861 GTTAT 2920

QY 1112 AAAATGAT 1169  
DB 2921 GTTAT 2980

QY 1170 AGAAATTTAGTACAAAGAGTACCAAAATATCTCTTATTTTAAATTTTACCAATTA 1229  
DB 2981 ATTAAGAT 3040

QY 1230 TAAAAATATCTCTTATTTTAAATTTTACCAATATATATATATATATATATATATATAT 1289  
DB 3041 GTTAT 3100

QY 1290 AATACCAACAAT 1349  
DB 3101 ATTAT 3160

QY 1390 ATATATCTGAT 1407  
DB 3161 AT 3220

QY 1408 CTATATCTGTCGCAACTAGTAAATCAATATATATATATATATATATATATATATATAT 1467  
DB 3221 ATTAAGTAT 3280

QY 1468 TTGAACATTTGTTTACCTGTTGATATTAAGATATATAAATCAATCAATATATATG 1527  
DB 3281 GTAGAAAT 3340

QY 1528 ACCATCTATTAAT 1557  
DB 3341 TATATGAATTAAGGTTTAAATTTTATTTT 3370

RESULT 12

ABL32254  
ID ABL32254 standard; DNA; 18154 BP.

XX ABL32254;

XX 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 227.

XX Human; immune system disease; cytosine methylation; antiasthmatic;

KM antiarteriosclerotic; antianaemic; cyostatic; noctropic;

KM neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KM antineumatic; antiarthritic; antidiabetic; antipsoriatic;

KM antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

KM acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

KM neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

XX gene; ds.

OS Homo sapiens.

PN WO200200928-A2.

PD 03-JAN-2002.

XX 02-JUL-2001; 2001MO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

PA Olek A, Piepenbrock C, Berlin K;

PI WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful

PT for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation -

XX PS Claim 1; SEQ ID NO 227; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

XX SQ Sequence 18154 BP; 5274 A; 112 C; 3279 G; 9489 T; 0 other;

XX Query Match 4.3%; Score 76.4; DB 24; Length 18154;

XX Best Local Similarity 47.2%; Pred. No. 0.0003;

XX Matches 303; Conservative 0; Mismatches 331; Indels 8; Gaps 2;

QY 927 TATATTAAGAAAATTTGTAATCATCATATATATATATATATATATATATATATATATAT 986  
DB 13819 TTTTATTAATTAATATTTTGTATATATATATATATATATATATATATATATATATAT 13878

QY 987 AAGTGTATCATTAAGATTTGAGAAACCAATATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1046  
DB 13879 TAAATTTAT 13937



QY	1047	TTTTCTAGTACTTTTCTTCAAGCCATATATAAACTTGTGAATGCTAAATGTATGCT	1106
Db	13938	TATTTTTTTTATTTTAAATAGATTTGTGTATATATTTTTTTTTTAAATGTTTTAT	13997
QY	1107	GGAAAAAATGTGATATGATATTCATATGAAATTTGTGATTTTCAAAGTCCAAAATCCATC	1166
Db	13998	TGATTTGTATGTTTTATGTTTTTAAAAAATATTTTGGATGTATATTTTTTTTATG	14051
QY	1167	AATGAAATTTAGTACAAAAGTACGCAAAATATTCCTATTTTAAATTTTACACA	1226
Db	14058	TTTTTATTTTATTTGTTTTTTTATTTTAAATAATTTGTATTTTAAAGTTTTG----	14112
QY	1227	ATATAAAAATTCCTATATTTTAAATTTTACAAATATATAATTTACCTGTACCTT	1286
Db	14113	--ATTAGAAATTAATTTATTTATTTAAATTTTTTAAAGTATATTTATTTTAA	14170
QY	1287	TAGAAATCCACCAATATTAATTAATCTAGATTTTATCTTAATTAATTTGAGATCTC	1346
Db	14171	TAAATTTTGAAGTATTTTTTTTATGTTAGATTTTATAGATTTTATATATTTATTAAT	14230
QY	1347	TCAATATATCGAATTTATTTTATATTTTGGACATATTTTCTATGTTTGAGTTAC	1406
Db	14231	TTTAAATATTTTATTTTATTTTGTATGATTTATTTTATTTATTTATTTTATTTGTATGT	14290
QY	1407	CCTTATATCTTGTGCAAACTAGTATATTCAAATATATGATTTTGAAGAACAATGCAT	1466
Db	14291	ATTTTAAAGTTTGAATATATATATATTTTATAGATAATATGATGTGTTTTATAGTG	14350
QY	1467	CTGAAAAATGTGTTTAACTGTTGTGGAATGTTAAAGTAATAAACAATGAGATAT	1526
Db	14351	TTTATATTTTATGTTTTATTTATTTATGTATTTGTGCAATATATTTTAAAGATATADA	14410
QY	1527	GACCATCTATTAATATCTCTCTTGTCTTTTAAAAAGTGT	1568
Db	14411	TAAATATGAAAAATATATAGTTTTTATTTATTTGTAATTTTTT	14452

```

RESULT 13
ABZ10015
ID ABZ10015 standard; DNA; 19634 BP

```

XX	AB210015;
XX	
XX	16-JAN-2003 (first entry)
XX	
XX	Haematopoietic cell proliferation disorder related DNA sequence #155.
XX	
XX	Human; haematopoietic cell proliferation disorder; cytostatic;
XX	gene therapy; lymphocytic leukaemia; acute myelogenous leukaemia;
XX	cytosine methylation state; gene; ds.
XX	
XX	Homo sapiens.
XX	
XX	WC0200277272-A2.
XX	
XX	03-OCT-2002.
XX	
XX	26-MAR-2002; 2002W0-EP03401.
XX	
XX	26-MAR-2001; 2001US-278333P.
XX	
XX	(EPIC-) EPICENOMICS AG.
XX	
XX	Berlin K, Braun A, Distler J, Guetig D, Howe A, Mueller J;
XX	Olek A, Piepenbrock C, Adorjan P, Grabs G, Lesche R, Lau E;
XX	Lewin A, Lipscher E, Maier S, Model F, Mueller V, Otto T;
XX	Pellet C, Schwepe I, Ziebarth H;
XX	
XX	WPI; 2003-018942/01.
XX	
XX	Detecting and differentiating between hematopoietic cell proliferative
XX	disorders, comprises contacting a target nucleic acid with a reagent

P1 that distinguishes between methylated and non-methylated CpG  
P2 dinucleotides -  
P3  
P4  
P5 Claim 28, SEQ ID 155, 117bp, English.  
P6  
P7  
P8  
P9  
P10  
P11  
P12  
P13  
P14  
P15  
P16  
P17  
P18  
P19  
P20  
P21  
P22  
P23  
P24  
P25  
P26  
P27  
P28  
P29  
P30  
P31  
P32  
P33  
P34  
P35  
P36  
P37  
P38  
P39  
P40  
P41  
P42  
P43  
P44  
P45  
P46  
P47  
P48  
P49  
P50  
P51  
P52  
P53  
P54  
P55  
P56  
P57  
P58  
P59  
P60  
P61  
P62  
P63  
P64  
P65  
P66  
P67  
P68  
P69  
P70  
P71  
P72  
P73  
P74  
P75  
P76  
P77  
P78  
P79  
P80  
P81  
P82  
P83  
P84  
P85  
P86  
P87  
P88  
P89  
P90  
P91  
P92  
P93  
P94  
P95  
P96  
P97  
P98  
P99  
P100

CC The present invention describes a method for detecting and  
CC differentiating between haematopoietic cell proliferative disorders  
CC associated with at least 1 gene and/or their regulatory regions in a  
CC subject. The method comprises contacting a target nucleic acid in a  
CC biological sample obtained from the subject with at least 1 reagent,  
CC which distinguishes between methylated and non-methylated CpG  
CC dinucleotides within the target nucleic acid, AB209661 to AB21118  
CC represent specifically claimed nucleotide sequences from the present  
CC invention. Oligonucleotides from the present invention can be used: for  
CC differentiating between healthy haematopoietic cells and proliferative  
CC disorder haematopoietic cells; for differentiating between acute  
CC lymphocytic leukaemia and acute myelogenous leukaemia; as probes for  
CC determining the cytosine methylation state and/or single nucleotide  
CC polymorphisms (SNPs) of haematopoietic cell proliferation disorder  
CC related sequences and their complements; and as primers for the  
CC amplification of haematopoietic cell proliferation disorder related  
CC DNA sequences. The nucleotide sequences from the present invention can  
CC also be used for detecting a predisposition to, differentiation between  
CC subclasses, diagnosis, prognosis, treatment and/or monitoring of  
CC haematopoietic cell proliferative disorders. The present method enables  
CC a highly specific classification of haematopoietic cell proliferative  
CC disorders allowing for improved and informed treatment of patients.

Sequence 19634 BP; 5528 A; 272 C; 3979 G; 9855 T; 0 other;

Query Match 4.3%; Score 76.2; DB 25; Length 19634;

Best Locally 45.48; Fied. NO. 0.00033;  
Matches 314; Conservative 0; Mismatches 373; Indels 4; Gaps 1

874 TTAAGGGGAGTGTAAATTATGTGTGTGAACCACCACTACCTTAGTAAGTATTATA 933

Db 6242 TTAATGTTTATGTTTATATTGTTGATTAGATTAAATTGTAAGAATAAAGTTTATT 6301

OY 934 AGAAATTGTAATCATCACATTATAATTGTCCTATTAAATTTATGATAAGTTGT 993

Db  
6302 TAGTGAGTGTCTTTTCGTTTTGGTATTGA AAAATTTT TTTT TTTTA 6361

994 ATCATTAGATTGAGAAACCAATAGTCTCTGTTGATTTTGAATTAATTCCTTCTA 1053

6362 ATTTATTAGATAGGAATAATAAATTCCTAATAATTATTTCTTATATATGAAATCTT 6431

1054 TGTACCTTCTTCAGGCTATATAAAAACTTTGTAATGCTAAATTCCTATGCTGCAAAZ 1111

[illegible]

Ox  
1114 AATGTTGAATTCAATACTTTCCTTTTTCCAAACAGCCTCGAATGG

[illegible]

Ov 1174 ATTATAGTCACAAACCTGAGCTCAGAGATTATTTTCCCTCTCTGGTGCGAATCATCACTGCTGTGC

100  
 90  
 80  
 70  
 60  
 50  
 40  
 30  
 20  
 10  
 0

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

0 20 40 60 80 100

[illegible][illegible]

1. The first step in the process is to identify the problem or issue that needs to be addressed. This involves gathering information and understanding the context of the problem.

0  
1  
2  
3  
4  
5  
6  
7  
8  
9  
A  
B  
C  
D  
E  
F  
G  
H  
I  
J  
K  
L  
M  
N  
O  
P  
Q  
R  
S  
T  
U  
V  
W  
X  
Y  
Z

[illegible]

QY 1470 GAAACATTGGTTTAACTGTGTGAATGTAAAGGTAAATAAACATTTCAGATTATGAC 1529  
 DB 6842 TTATAGTAAGATGTTTAAATTTTGTGTTTATTTTAAATTTTGTAGTTAATTT 6901  
 QY 1530 CATCTATTAATATACCTTCTTGTCTTTAA 1560  
 DB 6902 TAAATTTTATTTTAAATGATTATGATGAA 6932

## RESULT 14

ABL34123/c  
 ID ABL34123 standard; DNA; 10329 BP.

AC ABL34123;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 2096.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 XX antiarteriosclerotic; antianaemic; cyostatic; noctropic;  
 XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 XX gene; ds.

OS Homo sapiens.

XX WO200200928-A2.

PD 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

DR WPI; 2002-130909/17.

PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -

XX Claim 1; SEQ ID NO 2096; 32pp + Sequence listing; German.

XX The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.

XX Sequence 10329 BP; 2681 A; 56 C; 1865 G; 5727 T; 0 other;

Query Match 4.2%; Score 74.2; DB 24; Length 10329;

Best Local Similarity 47.8%; Pred. No. 0.00069;

Matches 313; Conservative 0; Mismatches 333; Indels 9; Gaps 3;

QY 712 TTACATTAACACACCTTCTTGAAGAGATTGAGAGAAATGTCAGTTACTT 771

DB 5962 TTCAATTTTAAACCTAATAATGACATCTTATATACATTATATTAATAACAC 5923

QY 772 TCTGAGTTCAATAGAGCACTTACAGACCTTTACTAA---ATACTCAAAAGAGA 827

DB 5922 TATCAATTCATATTTAATCATATATTCCTATTAATAAACATTCATCTCTTAATTA 5863  
 QY 828 AGATTTTAAACACTTAAGAAAGTAATGGAGTTAAAGAGCAACACTTAAGGGGAGGT 887  
 DB 5862 AAAAAATTAATAAAAAAACAACAAAAATTAACAAAAATTAATAAAAAAATTAACACA 5803  
 QY 888 TAAATTAATGTTGTGTAACACACCTACCTTTAGTAATATATTAAGAAATTTGTAATC 947  
 DB 5802 TTATTTTATTAACCTACCACTTAAATTCAAAATTAATAAAAAACAACAAACATA 5743  
 QY 948 ATCACTATTAATTAATTTGCTCTTATTTAAATTAATGTAAGTTGTAATGATGA 1007  
 DB 5742 ACCCTATTTACAAAAAATCTAATTTTAAATTAATTAACATTTCAAAAATA 5683  
 QY 1008 GAAACCAAAATAGCTGCTGTGATTTTGAATTAATTTGTTTCTATGTACTTTCTTC 1067  
 DB 5682 TATTAATTAATTAATCTCTAATTTTAAATTAATTAATCTTCAATTTAAATTAATTA 5623  
 QY 1068 AAGCTATATAAAAACTTTGTAAAT-GCTAAATTTGTAGCTGGAATAAAATGTAATGA 1126  
 DB 5622 AAACATAAAAAATTAACCTTTTCATTAACAACTAATTAATTAATCAACAAACCTTTTCATATA 5563  
 QY 1127 TTCAATTAATTAATTTGATTTTCAAGTCCAAATTCATCAATTAATTAATTAATTAATTA 1186  
 DB 5562 TTAATAAAAAATCAATTTTATCTCTATCAATCAACATCTTAATTAATTAATTAATAA 5503  
 QY 1187 CGTAACCTAAAAATATCTCTTATTTAAATTTTAAATTAACAAATTAATTAATTTCTTAT 1246  
 DB 5502 CAATATCTCAATTAATTTCCCTAAAAAATA-----AAACAAACTTAATAATTAATTAATA 5447  
 QY 1247 TTTAATTTTCAATTAATTAATTTATTCACCTGTACCTTTAGATTCACCAATATAT 1306  
 DB 5446 AATAAAAAATTAACCTAATAACAAATTAATCTTAATTTTTCACCAATCAATTAATCTTT 5387  
 QY 1307 TAACTTAGATATTTTATTTCTTAATTAATTTGAGATCTCAATTAATTAATCTGATA 1361  
 DB 5386 ATATATTTCCCAAAAAAACAATCAAAAAAATAATTAATCCATTTAAACAAACA 5332

## RESULT 15

ABL33013  
 ID ABL33013 standard; DNA; 7597 BP.

XX ABL33013;

DT 26-MAR-2002 (first entry)

DE Human immune system associated gene SEQ ID NO: 986.

XX Human; immune system disease; cytosine methylation; antiasthmatic;  
 XX antiarteriosclerotic; antianaemic; cyostatic; noctropic;  
 XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 XX antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 XX gene; ds.

OS Homo sapiens.

XX WO200200928-A2.

PD 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

PR 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;



**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 13:13:48 / Search time 4396 Seconds

(without alignments)  
16592.781 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783  
Sequence: 1 atccaactctgacatcttga.....tagagcgatcaagctgaacc 1783

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

GenEmbl:  
1: gb\_ba:\*  
2: gb\_hlg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_mu:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vi:\*  
30: em\_hlg\_hum:\*  
31: em\_hlg\_in:\*  
32: em\_hlg\_other:\*  
33: em\_hlg\_mus:\*  
34: em\_hlg\_pln:\*  
35: em\_hlg\_rod:\*  
36: em\_hlg\_mam:\*  
37: em\_hlg\_vit:\*  
38: em\_sy:\*  
39: em\_hlgo\_hum:\*  
40: em\_hlgo\_mus:\*  
41: em\_hlgo\_other:\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	206	11.6	2282	8	VFA277287
C 2	101.4	5.7	4601	3	DMU11584
C 3	101.4	5.7	19517	3	DMU37541
C 4	94.2	5.3	258658	3	AE014832
C 5	93.6	5.2	137889	9	AC073269
C 6	91.8	5.1	8056	6	AX599046
C 7	91.4	5.1	2009	6	AX457067
C 8	91.4	5.1	8056	6	AX599046
C 9	90.8	5.1	155106	9	AC104069
C 10	90.2	5.1	250029	3	AE014839
C 11	90.2	5.1	176368	9	CNS01RCG
C 12	90.2	5.1	181976	9	AC131649
C 13	90	5.0	104853	9	AC117444
C 14	90	5.0	150236	2	BX248102
C 15	90	5.0	153326	2	BX119991
C 16	89.8	5.0	108902	2	AC011430
C 17	89.6	5.0	118642	9	AC126283
C 18	89.6	5.0	130044	2	AC137080
C 19	89.4	5.0	30157	9	AC073242
C 20	89.4	5.0	94384	9	AC011718
C 21	89.4	5.0	106258	9	ALB07813
C 22	89.4	5.0	348600	1	AB063521
C 23	89.2	5.0	139409	9	AC069181
C 24	89.2	5.0	210797	9	AC013391
C 25	89	5.0	8056	6	AX598900
C 26	89	5.0	118642	9	AC126283
C 27	88.8	5.0	67870	3	PFMA113
C 28	88.8	5.0	152359	9	AC093790
C 29	88.8	5.0	159942	9	AC025018
C 30	88.4	5.0	104992	2	AC005504
C 31	88.4	5.0	138504	2	AC092665
C 32	88.4	5.0	159904	2	BX322565
C 33	88.4	5.0	169546	2	AC004157
C 34	88.4	5.0	250621	3	AE014849
C 35	88.2	4.9	349980	6	AX344567
C 36	88.2	4.9	1121	3	AF489467
C 37	88.2	4.9	14867	3	AE001398
C 38	87.8	4.9	349751	3	PFMA1493
C 39	87.2	4.9	141275	2	BX510640
C 40	87.2	4.9	313050	3	PF8929352
C 41	87	4.9	30157	9	AC073242
C 42	86.6	4.9	76568	3	WEREV
C 43	86.6	4.9	152966	9	AL359713
C 44	85.8	4.8	104853	9	AC117444
C 45	85.8	4.8	110000	2	PFMA113_03

#### ALIGNMENTS

RESULT 1  
VFA277287/c 2282 bp DNA linear PLN 07-DEC-2000  
LOCUS VFA277287  
DEFINITION *Vicia faba* partial *enodis* gene, exon 1 and 5' UTR.  
ACCESSION AJ277287  
VERSION AJ277287.1 GI:11602752  
KEYWORDS  
SOURCE  
ORGANISM  
*Vicia faba*  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae;  
*Vicia*.  
REFERENCE  
AUTHORS  
Hohnjec,N., Kuster,H., Albus,U., Frosch,S.C., Becker,J.D.,



```

Db      2805 TATAGTTTTTTTTTAAAAAAAAAATTTTATTTTTTAAAAAATTTTTTTTAAAAATGAAA 2746
QY      914 TACCTTAGTAGATATATAGAAAATTTGTAATCATACATTAATTAATTTGACCTATT 973
Db      2745 AATTAATTAATTAATTAATTTTCAATTAATAAATTTATTTATTTAAATTTTGTATTTTT 2686
QY      974 TAAATTAATTAATTAATTTGATCATTAAGATGAGAAACCAATAGCTCCTCTGAT 1033
Db      2685 AAAAAACATGATTTTATTAATAAATAATTTTTTAAAAAATTAACATTTAAGAAAATTT 2626
QY      1034 TTTTGATTAATTTGTTTCTATAGTACTTTCTTCAACCCATTAATAAAACCTTGTAATGC 1093
Db      2625 TAAAAATTTTATTAATTAATTTTAAAAATTTTAAATTTTCTATTAATAATTAATTA 2566
QY      1094 TAAATTTGATGCGGAAAAAATGTAATGAATTCATAGAAATTAATGATTTTCAAG 1153
Db      2565 TAAATATTCATATAATATATAATTTAATAAATAATTAATTAATTAATTAATTA 2506
QY      1154 TCCAAATTCATCAATTAAGAAATTTAGTCAAAACGTACTCAAAAATTTCTTATTT 1213
Db      2505 AAAAAAATTAATTAATTTTATTTTATTAATTAATTAATTAATTAATTAATTTAT 2446
QY      1214 AAATTTTCAACAATATTAATAAATTTCTCTTATTTTAAATTTTACATTAATTAATTTAT 1273
Db      2445 TTAATTTATGTTATTAATTAATTTATTTATTTATTAATTAATTAATTAATTAATTT 2386
QY      1274 CACCTGTCACCTTAGAATTCACCAACAATTAATTAATTAATTAATTTATTTCTTA 1333
Db      2385 ATTTAATTAATTAATTTATTAATTTATTTATTTATTTAAAAATTTATTTTCAATTTA 2326
QY      1334 ATTTTGATCTCTCAATATATCGATATTTATTTATTTATTTGTCATTTTCTTTAG 1393
Db      2325 ATATATATTAATTAATTAATTAATTTATTTATTTATTTATTTATTTTAA--ATAATTTATTTA 2268
QY      1394 TTTTAGATTAACCTTATATCTTGGTCAACAGTATTAATTAATTAATTAATTAATTTGTAAG 1453
Db      2267 TAAATATATTTATTAATAAATAGTTTATTAAGATTAATTAATAAATAATTTTTTTTA 2208
QY      1454 GACGATTCGATCTTAACATTTGTTTAACTGTTGGAATGTAAGTAATPA 1513
Db      2207 AAAAAAATTAATTTTAAATTTTAAATTAATTAATAAATTTAAGATAGGGGAAATPA 2148
QY      1514 CATTCAGATTAATGACCATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1573
Db      2147 TTAATTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 2086
QY      1574 A 1574
Db      2087 A 2087

```

```

RESULT 3
DMU37541/c 19517 bp DNA circular INV 02-MAR-2001
LOCUS Drosophila melanogaster complete mitochondrial genome.
DEFINITION U37541
ACCESSION U37541
VERSION U37541.1 GI:1166529
KEYWORDS
SOURCE
ORGANISM
.
mitochondrion Drosophila melanogaster (fruit fly)
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Phyloroidae; Drosophilidae; Drosophila.
1 (bases 12511 to 12682)
Clary,D.O., Goddard,J.M., Martin,S.C., Faure,C.M. and
Wolstenholme,D.R.
REFERENCE
AUTHORS
TITLE
JOURNAL
Nucleic Acids Res. 10 (21), 6619-6637 (1982)
MEDLINE
PUBMED
6294611
2 (bases 5269 to 5695)
Clary,D.O., Wahleithner,J.A. and Wolstenholme,D.R.
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
63220794
Nucleic Acids Res. 11 (8), 2411-2425 (1983)
Genes
3 (bases 404 to 5272)
de Bruijn,M.H.
Drosophila melanogaster mitochondrial DNA, a novel organization and
genetic code
Nature 304 (5923), 234-241 (1983)
JOURNAL
MEDLINE
83245048
PUBMED
6408489
4 (bases 804 to 1778)
Satta,Y., Ishiwa,H. and Chigusa,S.I.
Analysis of nucleotide substitutions of mitochondrial DNAs in
Drosophila melanogaster and its sibling species
Mol. Biol. Evol. 4 (6), 638-650 (1987)
JOURNAL
MEDLINE
88174373
PUBMED
2832697
5 (bases 5268 to 13619)
Garesse,R.
Drosophila melanogaster mitochondrial DNA: gene organization and
evolutionary considerations
Genetics 118 (4), 649-663 (1988)
JOURNAL
MEDLINE
88212147
PUBMED
3130291
6 (bases 441 to 2967)
Satta,Y. and Takahata,N.
Evolution of Drosophila mitochondrial DNA and the history of the
melanogaster subgroup
Proc. Natl. Acad. Sci. U.S.A. 87 (24), 9558-9562 (1990)
JOURNAL
MEDLINE
91088557
PUBMED
2124697
7 (bases 14215 to 14512)
Ballard,J.W., Olsen,G.J., Fatch,D.P., Odgers,W.A., Rowell,D.M. and
Atkinson,P.W.
Evidence from 12S ribosomal RNA sequences that onychophorans are
modified arthropods
Science 258 (5086), 1345-1348 (1992)
JOURNAL
MEDLINE
93088057
PUBMED
1455227
8 (bases 14917 to 19517)
Lewis,D.L., Farr,C.L., Farquhar,A.L. and Kaguni,L.S.
Sequence, organization, and evolution of the A+T region of
Drosophila melanogaster mitochondrial DNA
Mol. Biol. Evol. 11 (3), 523-538 (1994)
JOURNAL
MEDLINE
94285822
PUBMED
8015445
9 (bases 1 to 408; 13319 to 19517)
Lewis,D.L., Farr,C.L. and Kaguni,L.S.
Drosophila melanogaster mitochondrial DNA: completion of the
nucleotide sequence and evolutionary comparisons
Insect Mol. Biol. 4 (4), 263-278 (1995)
JOURNAL
MEDLINE
96423163
PUBMED
8825764
10 (bases 1 to 19517)
Lewis,D.L., Farr,C.L. and Kaguni,L.S.
Direct Submission
Submitted (03-OCT-1995) Laurie S. Kaguni, Biochemistry Department,
Michigan State University, East Lansing, MI 48824-1319, USA
location/Qualifiers
1. 19517
/organism="Drosophila melanogaster"
/organism="mitochondrion"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/note="derived from new and previously submitted
sequences; sequence is a composite containing sequences
obtained from different Drosophila melanogaster strains"
1. .65
/product="rRNA-11e"
complement(97. .165)
FEATURES
source

```

```

CDS
    /product="tRNA-Gln"
    /protein_id="AAC47811.1"
    /codon_start=1
    /transl_table=5
    /product="NADH dehydrogenase subunit 2"
    /protein_id="AAC47811.1"
    /db_xref="GI:1166530"
    /translation="MENNSSKILFTITMIGTILITWTNSGILAGMGLLEINILFPIPLSDNNKLNSTERSLSKYEPIQLQVLASTVLLPSILILMKNNKNNENINSEFSTNMSALLLKSGNAPHPHFPPNNKEGLTWNNALMLMTQKIRAPMLNLSLNKILKLLSVLTSVILGAIQGLNQTSLRKLMAFSSINHLGMLNLSLSSIMILFFVFSISFVLTFNENFKLFIENLQNFSSNFVNSKILKFTLFNNFLSLGGLPEPILGPKMLVIOQLTLCNQFMPTIMMSTLTLLFFYLIRICYSAPMNMYFENNIMKKNMNSINYNMIMTFFSIFGLFPISLFFYFMF"
    1264..1329
    /product="tRNA-Tip"
    complement(1322..1383)
    /product="tRNA-Cys"
    complement(1403..1468)
    /product="tRNA-Tyr"
    join(1470..1472,1474..3009)
    /note="mechanism underlying reading frame shift after first codon uncertain"
    /codon_start=1
    /transl_table=5
    /product="cytochrome c oxidase subunit I"
    /protein_id="AAC47812.2"
    /db_xref="GI:7412849"
    /translation="MSRQMLFSTNNKIDIGTLTYFGAWAGWVGTSILLRAELGHPGLALIGDQITNIVATMAHATIMFFMYMPTMIGFEGNNVYELMLGAPDAFPRNNMSEWLLPPLSLILSVSWENAGAGTGVTPPLSLGILHAGASVDLAFSLHLGISLTLAAVNEFTIVINMRSTGISLDRMPLEFWSVITALLLSLPLVLAGAITMLLTDRLNITSEFPDAPGSGEPIQLQHLFFEFHPEVYILLPFGMISHSIQSGSKETFGSLGNTVMYLAIGLGFVMAHMEFTVGMDTRAYTSATMIIAVPTGKIFSMATLGLQGLSYSPLAALGAEPLFTVGLGVLAVLANSVDIILHDPYVVAHFVLSGWAFAIMAEAFIMHYPLFQCLTINNMKLSHPIIMVIGVNLFFPDHGLGAGMRGRYSDDPATYNNIYSTVIGTSLSLGILFFFTIWESIVSRQVIYIQLNLSIEYQNPRAHSISELPLLTN"
    3012..3077
    /product="tRNA-Leu"
    3083..3167
    /note="TAA stop codon is completed by the addition of 3' A residues to the mRNA"
    /codon_start=1
    /transl_except=(pos:3167,aa:TERM)
    /transl_table=5
    /product="cytochrome c oxidase subunit II"
    /protein_id="AAC47813.1"
    /db_xref="GI:1166532"
    /translation="WSTNATGLODASPLMEQLFFPHDALITVMITVYGLMEMLFFNNVNSFELHAGQILEMTITLPAITLILFALPSLILVLDLEINSEPTVLSIGHQWVWSYSEDFENNIEFDSYMTPTNELMTDGFLLDVNRVYLPNNSQIRLLIVTADYVHMTVPLALCVKQDTPGRINQTFPIINRGLFYGCSEBICGANSFMEIVTESVPMVYFIKWSISNNS"
    3168..3388
    /product="tRNA-Lys"
    3840..3906
    /product="tRNA-Asp"
    3907..4068
    /codon_start=1
    /transl_table=5
    /product="tRNA-Pase 8"
    /protein_id="AAC47814.1"
    /db_xref="GI:1166533"
    /translation="MPQWAPISLWLLPITISITPLIFCSINITYSTMPSPSNELKAINNNSNMWKM"
    4062..4736
    /codon_start=1
    /transl_table=5
    /product="tRNA-Pase 6"
    /protein_id="AAC47815.1"

```

CDS		/db xref="GI:116534"	
		/translation="MTNLSVDFDLAFNFSLMNLSTFLGLMLTPISYLWLPKRVNI MNSLTLLTHKEFKTLIGSGHNGSTPFIISFSLIFNNMGLEPYIFETSHLTLT LSALPIMLCFYLXGMIHHOMGFHALPOGPATLMFMVCIENTISMISRGTAAR LTANITAGHLTLTLTGNTGSSMYLMTPLFMAQIALLVBSAVAMIOSYPAVLST YSSEVN"	
	4736..5524	/codon start=1	
		/transl table=5	
		/product="cytochrome c oxidase subunit III"	
		/protein_id="AAC47816.1"	
		/db xref="GI:116535"	
tRNA		/translation="MSTHSNHPEPHLDVSPEDLTGAIGATTVSCKPFHQDYDLSF VGNIITLITLYVMWRDVSRSBGTYOGLHYAATGLRGMILLFILSELFPVSWFAF FHSSSIPALIEASMPMGLISENFQJPLANTALIVSGVTPVAHNSHMENNSQT TGGLEPTVLIGIYFTILOAYEIEADPTIADISYSTFPMATGVFGIHVLIIGTEPLL CILRHNNRSGSNHHRGFEEAAWMHFVDVWLFLYILTVMGC"	
	5543..5607		
		/product="tRNA-Gly"	
CDS		/codon start=1	
	5608..5961	/transl table=5	
		/product="NADH dehydrogenase subunit 3"	
		/protein_id="AAC47817.1"	
		/db xref="GI:116536"	
		/translation="MESIIFFALLILTTITYMFASLSISKALIDREKSSPEEGFD PXSSSLPSRLRFPLITIIFFLDEVALIHPMITIMKSYMIMTIYTIIIFILILI GLYHEMNQCLNMNSN"	
tRNA		5981..6045	
		/product="tRNA-Ala"	
tRNA		6055..6118	
		/product="tRNA-Arg"	
tRNA		6119..6183	
		/product="tRNA-Asn"	
tRNA		6184..6251	
Query Match	5.7%; Score 101.4; DB 3; Length 19517;		
Best Local Similarity	46.5%; Pred. No. 2.4e-06;		
Matches	365; Conservative 0; Mismatches 416; Indels 2; Gaps 1;		
OY	794 TACGACACTTCTACTCAATAATACACAAAGAGAGACATTTTAAACACTTAGAGAAGTAAT	853	
Db	17781 TAATATATATATATATTAATTAATTAATCATATATATATATATATATATATAGTA	17722	
OY	854 GGAGGTAAAGACACACATTTAAGGGAGAGCTTAAATTAATGCTGTGAACACCAC	913	
Db	17721 TATAGTTTTTTTAAAAAAAATTAATTTTTTTTAAAAAATTTTTTTTAAAAATGAAA	17662	
OY	914 TACCTTAGTAGTATTAATAGAAATGTTAAATCATCACATTAATATATATTCCTTAT	973	
Db	17661 AATTAATTAATTTATTTTCATTTAAATTTATTTATTTAAAAATTTTTTGTTAATTTT	17602	
OY	974 TAAAAATTGATAAAGTTGTATCATTTAAGATTGAGAAAACCAAATAGTCTGTGAT	1033	
Db	17601 AAAAACATGATTTTATATATTAATTAATTTTTTAAAAATATATACATTTAAGAAAATTT	17542	
OY	1034 TTTTGAATTAATGTTTTTCATGATGTACTCTTTCTCAAGCCTATATAAAAACCTTGTATGC	1093	
Db	17541 TAAAAAATTTATATTAATTTATTTAAATTAATTTATTTTCATATATATATATATATTA	17488	
OY	1094 TAAATTTGATCGGAAAAAATGTGTAAATGAATTCATATGAATTAATGATTTCCAAG	1153	
Db	17481 TAATATATCAATATATATATTAATTTTAAAAATATATATATTAATTAATTAATTA	17422	
OY	1154 TCCAAAATCATATATGAATTTTACTACAAACGTAACGCCAAAATATATTCCTTATTT	1213	
Db	17421 AAAAAAAAAAAAAAAAAATTTTTTATATTAATTAATAATATAGTAATTAATTTTAT	17366	
OY	1214 AAATTTTACACAATATAAAAAATTTCTGTTATTTTAAATTTTACAAATATATATTTAT	1273	
Db	17361 TTATTTTATAGTTATTTAAATTTTATTTATTTATTTAAATATATTAACAAGAATGATTT	17300	
OY	1274 CACCTGTACCTTTAGATATCCACCAACAATATTAATATCTTAGATATTTTATTTCTTAATA	1333	





repeat_region	complement (3821, .2848)	/db_xref="GI:23495048"
repeat_region	/note="AT_rich"	/translition="WRIFGSLKRMVYTPYEYTYLLCNEQSKSRKRFNPFSTMTYNTSEL
repeat_region	/rpt_type=tandem	ONNMKNLENNMGLNLTISYGIKKSVKYGYNKMFPHYNLLKFCYCNTPEDDIIILLO
repeat_region	complement (3335, .3365)	SVALKERREFEVYLSLNKLVIIFKDMNEQEVNQVYNIYKYLIAEQUNASDVEII
repeat_region	/note="AT_rich"	TMFLIKIKNHLIYGLKKSIIILHSLTKLKIYDQDLDPISYIVNNEHKMCAVSH
repeat_region	/rpt_type=tandem	ITSAVSKOINIKFLOLYKYLLOIYENIKYMSLSIYTLIOIKPILOTVDNYSLY
repeat_region	/rpt_unit=(A)n	KTODYCTLOFNKNSGSSPEDNLEOKKKCKDYDKNIYNESNVCCKYKEVSLN
repeat_region	complement (4042, .4111)	NONSNNNTTSHSVKVTHTYQSKETYPHTIPLVDYERKTKCTTNNINNDKROCK
repeat_region	/note="AT_rich"	NDENNVNIFSNSSKHSIISNDEYCHADLTAKINNKEQNNLNKKKTLNIOIPLLS
repeat_region	/rpt_type=tandem	KVNNLALSLFQYKLACTYKELNYFNITVYKSLPYLAKIOQNKISPECTYLL
repeat_region	complement (3861, .3880)	BCFTLVPIDNNIDELIINILTNLETVTFEYHNYLCRILOCFQHLTIYDNLTSIC
repeat_region	/rpt_type=tandem	IIIFKKKAFPERCNLKDLOLFLHFYEKNSDEQEMINYLKQIIEONNHIILOSBNIS
repeat_region	/note="AT_rich"	OODTHQHSNTTNNMNIYNDKXLIIVKYNKQCFNEKEKNIYEATHTVHDKDK
repeat_region	/rpt_type=tandem	EMDGRPHNENNDYKKSVDYIFNTNIEQKNI"
repeat_region	complement (3453, .3531)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (3661, .3880)	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(A)n	
repeat_region	complement (4042, .4111)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (4275, .4301)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (4348, .4369)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (4794, .4992)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (4833, .4872)	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(TA)n	
repeat_region	4837, .4872	
repeat_region	/rpt_type=tandem	
repeat_region	5124, .5152	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(A)n	
repeat_region	complement (5154, .5178)	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(TA)n	
repeat_region	5275, .5331	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(A)n	
repeat_region	5280, .5323	
repeat_region	/rpt_type=tandem	
repeat_region	complement (5332, .5357)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (5378, .5402)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (5451, .5484)	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(TA)n	
repeat_region	complement (5495, .5532)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (5578, .5604)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (5670, .5689)	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(A)n	
repeat_region	complement (5749, .5838)	

repeat_region	complement (3335, .3365)	/db_xref="GI:23495048"
repeat_region	/note="AT_rich"	/translition="WRIFGSLKRMVYTPYEYTYLLCNEQSKSRKRFNPFSTMTYNTSEL
repeat_region	/rpt_type=tandem	ONNMKNLENNMGLNLTISYGIKKSVKYGYNKMFPHYNLLKFCYCNTPEDDIIILLO
repeat_region	complement (3335, .3365)	SVALKERREFEVYLSLNKLVIIFKDMNEQEVNQVYNIYKYLIAEQUNASDVEII
repeat_region	/note="AT_rich"	TMFLIKIKNHLIYGLKKSIIILHSLTKLKIYDQDLDPISYIVNNEHKMCAVSH
repeat_region	/rpt_type=tandem	ITSAVSKOINIKFLOLYKYLLOIYENIKYMSLSIYTLIOIKPILOTVDNYSLY
repeat_region	/rpt_unit=(A)n	KTODYCTLOFNKNSGSSPEDNLEOKKKCKDYDKNIYNESNVCCKYKEVSLN
repeat_region	complement (4042, .4111)	NONSNNNTTSHSVKVTHTYQSKETYPHTIPLVDYERKTKCTTNNINNDKROCK
repeat_region	/note="AT_rich"	NDENNVNIFSNSSKHSIISNDEYCHADLTAKINNKEQNNLNKKKTLNIOIPLLS
repeat_region	/rpt_type=tandem	KVNNLALSLFQYKLACTYKELNYFNITVYKSLPYLAKIOQNKISPECTYLL
repeat_region	complement (3861, .3880)	BCFTLVPIDNNIDELIINILTNLETVTFEYHNYLCRILOCFQHLTIYDNLTSIC
repeat_region	/rpt_type=tandem	IIIFKKKAFPERCNLKDLOLFLHFYEKNSDEQEMINYLKQIIEONNHIILOSBNIS
repeat_region	/note="AT_rich"	OODTHQHSNTTNNMNIYNDKXLIIVKYNKQCFNEKEKNIYEATHTVHDKDK
repeat_region	/rpt_type=tandem	EMDGRPHNENNDYKKSVDYIFNTNIEQKNI"
repeat_region	complement (3453, .3531)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (3661, .3880)	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(A)n	
repeat_region	complement (4042, .4111)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (4275, .4301)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (4348, .4369)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (4794, .4992)	
repeat_region	/note="AT_rich"	
repeat_region	/rpt_type=tandem	
repeat_region	complement (4833, .4872)	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(TA)n	
repeat_region	4837, .4872	
repeat_region	/rpt_type=tandem	
repeat_region	5124, .5152	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(A)n	
repeat_region	complement (5154, .5178)	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(TA)n	
repeat_region	5275, .5331	
repeat_region	/rpt_type=tandem	
repeat_region	/rpt_unit=(A)n	
repeat_region	5280, .5323	
repeat_region	/rpt_type=tandem	
repeat_region	complement (5332, .5357)	

```

Db      186942 TTATTATTATT-----TATTATTATAATAATAATTATTGTTATTATTATTATT 186889
Cy      1318 TATTATTCTTATAATTGAGACTCTCAATATATCGATTTATTATTATTGTT 1377
Db      186888 TATTATTATTATTATTATTATTATTATAATAATAATTGTTATTATTATTATT 186829
Cy      1378 GTCAATTCTTCTAGTTAGAGTTAACCTATATCTGGCAACTGTAATCAAT 1437
Db      186828 TTTTATTATTATTATTATTATTATTATAATAATAATTATTGTTATTATTATT 186769
Cy      1438 ATATGAGTT 1446
Db      186768 TTTATTATT 186760

RESULT 5
LOCUS   AC073269 137889 bp DNA linear PRI 29-APR-2003
DEFINITION Homo sapiens BAC clone RP11-436F9 from 7, complete sequence.
ACCESSION AC073269
VERSION   AC073269.7 GI:14269824
KEYWORDS  HTG.
SOURCE    Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 137889)
  AUTHORS Sulston,J.E. and Waterston,R.
  TITLE   Toward a complete human genome sequence
  JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
  PUBMED  99063792
  MEDLINE 9847074
REFERENCE 2 (bases 1 to 137889)
  AUTHORS Abbott,A., Mclellan,M. and Vanhurnt,A.
  TITLE   The sequence of Homo sapiens BAC clone RP11-436F9
  JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 137889)
  AUTHORS Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL Submitted (12-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE 4 (bases 1 to 137889)
  AUTHORS Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL Submitted (31-MAY-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE 5 (bases 1 to 137889)
  AUTHORS Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL Submitted (01-JUN-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE 6 (bases 1 to 137889)
  AUTHORS Waterston,R.
  TITLE   Direct Submission
  JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 7 (bases 1 to 137889)
  AUTHORS Waterston,R.
  TITLE   Direct Submission
  JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 8 (bases 1 to 137889)
  AUTHORS Waterston,R.
  TITLE   Direct Submission
  JOURNAL Submitted (29-APR-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT  On May 31, 2001 this sequence version replaced gi:13794256.
            ----- Genome Center
            Center: Washington University Genome Sequencing Center

```

Center code: MUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_NH0436609

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send [mailto:egreen@nhgri.nih.gov](mailto:mailto:egreen@nhgri.nih.gov), or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E., Tateo,M., Catanese,J.V. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>  
 VECTOR: pBac3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-29C24, 2000 bp overlap the clone sequenced to the right is CTD-2016H9, 200 bp overlap. Actual start of this clone is at base position 1 of RP11-436F9 actual end is at base position 92636 of CTD-2016H9.

#### FEATURES

##### SOURCE

```

1..137889
  /organism="Homo sapiens"
  /mol_type="genomic DNA"
  /db_xref="taxon:9606"
  /chromosome="7"
  /map="7"
  /clone="RP11-436F9"
  /clone_lib="RPCT-11"
  453..773
    /rpt_family="L1"
    /rpt_family="L1"
    /rpt_family="Alu"
    1063..1316
      /rpt_family="L1"
      /rpt_family="L1"
      1351..1469
        /rpt_family="L1"
        /rpt_family="L1"
        2098..2256
          /rpt_family="MIR"
          2505..2753
            /rpt_family="L1"
            2971..3034
              /rpt_family="L1"
              3358..3742
                /rpt_family="L2"
                3905..4033
                  /rpt_family="L2"
                  4148..5347
                    repeat_region

```

Query Match	5.2%	Score 93.6;	DB 9;	Length 137889;
Best Local Similarity	49.0%	Pred. No. 2e-05;		
Matches 339;	Conservative 0;	Mismatches 344;	Indels 9;	Gaps 3;
OY	918	TTTAGTAAAGTATTAAGAATAATTGGTACATCATCA-TTA-TAATTANTGCTGTATTTAA	976	
Db	63049	TTTATATATATATATATATATATTTTAAATATATATATTTTATTCATCATTAATATATATTC	6310	
OY	977	AATATAGTAAAGTTTGATCCTTATAGATGAGAAACAAPAGTCCGTGCCTGTGATTT	1036	
Db	63109	TTAATATATATATATATATTTTAAATATATTTTATATATATATATATATATTTTAAATAT	6316	
OY	1037	TGAATTAATGTTTTCTATAGTTACTTCTTCTCAAGCCATATATAAAACCTTGGATGCTAA	1096	
Db	63169	ATATATATTTTAT	63222	
OY	1097	ATTGTATGCGGAAAAAANGTGAATGCAATTCGAATATGATATGATATGATATGAAAGTCC	1156	
Db	63229	ATATATATATTTTAAATATATATATACATTTTATATATATATATATATATATATATATTT	63288	
OY	1157	AAAAATCCATC----AAAGAAATTTAGTACAAACGTAACGTAACCAAAATATCTCTAT	1210	
Db	63289	ATTATATATATGATATATATATATTTTAAATATATATATATATATATATTTTAAATATAT	63346	
OY	1211	TTTAAATTTTACAAATATATAAAAATTCCTCTATTTTAAATTTTACAAATATATATAT	1270	
Db	63349	TATATATATATATTTTAAATATATATATTTTATATATATATATATATATATATATATAT	63400	
OY	1271	TATCACCTGTACCTTTAGAAATGCCAACAAATATATATATCTAGATATTTATTTCTTA	1330	
Db	63409	ATTATATATATATATATATATATATTTTAAATATATATATATATATATATATATATATA	63466	
OY	1331	ATAATTTTGAATCTCAATATATCTGATATATTTATTTATTTTGGTCATATTTTCTT	1390	
Db	63469	ATA--TTTATATTTTATTTAT	63522	
OY	1391	ATGTTTAAAGTAAACCCTTATATCTTGTCCAACAGTATATCAATATATGAGTTTGG	1450	
Db	63527	ATA	63586	
OY	1451	AAGACACATTTGACATCTTGAAACATTTGTTTAAACCTTGAGTAATGTTAAAGTAAATA	1510	
Db	63587	TATACGATATATATATATTTTAAACATATATATATATATATTTTAAACATATATATATA	63646	
OY	1511	AAACATTCAGAAATTATAGACATCTATATATATCTCCCTTTGCTTTTAAAAAAGTGGC	1570	
Db	63647	TATTTTAAATATTTTAAATATATATATATATATATATATATATTTTATATATATATAT	63706	
OY	1571	ATGAAATGCTCTATAGTATAGTAGTAGTGTCT	1602	
Db	63707	TTTAAATATATATATATTTTATTTATATATATAT	63738	
RESULT 5				
AX539046/c				
LOCUS	AX539046	Sequence 386 from Patent WO02077272.	DNA	linear PAT 14-FEB-2003
DEFINITION	AX539046			
ACCESSION	AX539046.1	GI:28399186		
VERSION				
KEYWORDS				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS	Berlin,X., Braun,A., Distler,J., Guetig,D., Howe,A., Mueller,J., Olek,A., Pilsenbrock,C., Adorjan,P., Grabs,G., Lesche,R., Iehl,E., Lewin,A., Lipscher,E., Maier,S., Model,F., Mueller,V., Octo,T., Pelzer,C. and Ziebarth,H.			
TITLE	Methods and nucleic acids for the analysis of hematopoietic cell proliferation disorders			
JOURNAL	Patent: WO 02077272-A 386 03-OCT-2002;			
FEATURES	Epigenomics AG (DE)			
	Location/Qualifiers			



Qy	799	AAACCTTAACTAAATATACACAAAGAAGGAAAGTTTAAACAATTGGAAGAATGAGAG	858
Db	626	AAATATTTTAAATAATTTTAAATAATTAATTTATTAATAAAAAAAAAAAAAAATATTA	685
Qy	859	TTAAAGACACACATTAAGGGGAGTGTAAATTAATGTGTGAACCACTACTACT	918
Db	686	AATATAAAAAATTTTATTTATTTAAATTTATTTAAATTAATTTATTTATTTAA	745
Qy	919	TTAGTAGATTTATTAAGAAATTTGATCATCACATTAAATATTGCTTATTTAA	978
Db	746	TTTTTAAATTTAAATTTAATTTAAATATATTAATTTATTTATTTATTTAAAT	805
Qy	979	TTATG-ATTAAGTGTATCATTAAGTTAGAAAAACAATAGCTCTGCTGATTTT	103
Db	806	AAATTATTAAGATTTATTAATAATTTTATTAATAAAAAATGTTTTTTTATTTTTT	865
Qy	1038	GATTATGTTTTTCATGTACTTTCTCAAGCTATATATAAAAACTTTGATGCTAA	1097
Db	866	TTATGTTAAAAAATTTGTAATTTATTTAAATTTAATTTTATTTATTTTCTGTTTT	925
Qy	1098	TGTATGCGGAAAAAATGTGAATGAATTCATATGAATATATGTTTCAAGTCA	1157
Db	926	GTTATTAATTTTTTTAAATAATTTTATTTTTTTATTTAAAGATTAATAATATA	985

Center: Washington University Genome Sequencing Center  
 Center code: NUSSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [sepiens@wustl.wustl.edu](mailto:sepiens@wustl.wustl.edu)  
 ----- Summary Statistics -----  
 Center project name: H\_NH0279B11  
 Drafting Center: WIBR

NOTICE: This sequence may not represent the entire insert of this

repeat_region	/rpt_family="AT_rich"	11208	.12077
repeat_region	/rpt_family="L1"	12114	.12173
repeat_region	/rpt_family="Alu"	12178	.13052
repeat_region	/rpt_family="ERV1"	13053	.13824
repeat_region	/rpt_family="ERV1"	13823	.14249
repeat_region	/rpt_family="ERV1"	14232	.14785
repeat_region	/rpt_family="ERV1"	14438	.14542
repeat_region	/rpt_family="CT-rich"	17655	.18541
repeat_region	/rpt_family="ERV1"	18542	.18780
repeat_region	/rpt_family="Alu"	18787	.18983
repeat_region	/rpt_family="L1"	19002	.19280
repeat_region	/rpt_family="Alu"	19654	.20155
repeat_region	/rpt_family="L1"	21441	.21600
repeat_region	/rpt_family="(TA)n"	21970	.21990
repeat_region	/rpt_family="AT_rich"	25855	.25609
repeat_region	/rpt_family="AT_rich"	26541	.26612
repeat_region	/rpt_family="L1"	26707	.26999
repeat_region	/rpt_family="Alu"	26971	.26999
repeat_region	/rpt_family="(A)n"	27033	.27055
repeat_region	/rpt_family="(CA)n"	27219	.27253
repeat_region	/rpt_family="(CA)n"	27428	.27813
repeat_region	/rpt_family="MeIR"	27883	.27922
repeat_region	/rpt_family="(TAGA)n"	27884	.28005
repeat_region	/rpt_family="L1"	27976	.28007
repeat_region	/rpt_family="AT_rich"	28740	.28798
repeat_region	/rpt_family="CT-rich"	30567	.30878
repeat_region	/rpt_family="Alu"	30849	.30882
repeat_region	/rpt_family="(A)n"	31141	.31392
repeat_region	/rpt_family="MIR"	31419	.31765
repeat_region	/rpt_family="Alu"	31708	.31741
repeat_region	/rpt_family="(CAAA)n"	31810	.31922
repeat_region	/rpt_family="MIR"	32437	.32533
repeat_region	/rpt_family="(TATATG)n"	32463	.32531
repeat_region	/rpt_family="ERV1"	32533	.32626
repeat_region	/rpt_family="MIR"	32662	.32966
repeat_region	/rpt_family="Alu"	32958	.33095
repeat_region	/rpt_family="GA-rich"		

Query Match	5.1%;	Score 90.8;	DB 9;	Length 155106;
Best Local Similarity	47.2%;	Pred. No. 5.1e-05;		
Matches 308;	Conservative	0;	Mismatches 342;	Indels 2;
				Gaps 1.

99218 AATATGATTTTATATATAATAATTTATATATATTTTATATATATATATATATA 99277

99338 ATTATATATAATATTATAAATAATTTTATATAATAATATATATATTT 99339

b  
99458 AATATATAATTAAATAATTTTATATATAATTAATATATAATTAATATAA 99517  
1311 ACTTAGATATTTATTCCTTAATAATTTTGAGCTCCTCAATATACGTGATTTATTTA 1370

[illegible]

99626 ATTTTATATATAATATATTAATATATATTTATATATATATATATATTT 9975  
1551 TGCTTTAAAAAGCTGCATGAAATGCTCTATGCTAGAGTGTCT 1602

LOCUS	250029 bp	DNA	linear	INV 07-OCT-2002
AE014839				
AE014839				
plasmidium falci				
parum 3D7				
chromosome 11				
section 4 of 8				
of the				
complete				
sequence.				

KEYWORDS	AE014639.1	GI:25456131
VERSION		
SOURCE		
ORGANISM	Plasmodium falciparum 3D7	
	Plasmodium falciparum 3D7	

1 (bases 1 to 250029)  
Gardner, M. J., Hall, N.,  
Fung, E., White, O., Berriman, M., Hyman, R. W.,  
Nelson, K. E., Bowman, S., Paulsen, I. T.

Genome sequence of the human malaria parasite *Plasmodium falciparum* Nature 419 (6906), 498-511 (2002)

12368864  
2 (bases 1 to 250029)

Gardner, M.J.  
Direct Submission

Submitted (13-SEP-2002) The Institute for Genomic Sciences and  
Medical Center Dr., Rockville, MD 20850, USA

Location/Qualifiers  
1. .250029

```
/organism="Plasmodium falciparum 3D7"
/mol_type="genomic DNA"
```

```
/isolate="3D7"  
/db_xref="taxon:
```

```

/chrchromosome="11"
region 325, .384

```

```

-      /rpt_type=tandem
region 556. .604

```

```

/ rpt_type=tandem
<960, .>2083

```

```
/locus_tag="PF11  
join(<960, .993,
```

```
/locus_tag="PF11
join(960.1 993.1
```

```
/locus_tag="PF11  
/codon_start=1
```

```
/product="phosph  
/protein_id="AAN
```

```

/db_xref="GI:234
/translation="MTT

```

7 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1

VLPEWFDHIAPDILAN  
DENI.KPIKHVYIJDSE

```

region
1404. 1438
/rot type= tandem

```

```

/1pc_cylpc=curadem
2137. .2163
/rot timot andem

```

```

region
2291.  .2349
/tpc_type=bandwidth
/tpc_type=bandwidth

```

```

region
2589, .2632
/tpc_type=candem
/mnt_time=bandem

```

```

/ rpl_type=candem
2864, .2894
/ next_time=1andot
region

```

```

region
3372. 3594
/ rpl_type=land

```

```

region: 3632. 3660
      /rpt_type=landfill

```

```

/ rpt_type=candem
3830.  .3862
region

```

```

/ rpt_type=candem
4056. .4099
region

```

```
/rpt_type=Candem
<4298, .>4483
```

```
/ locus_tag="PF11
<4298. .>4483
```

```
/locus_tag="pf11
4298. .4483
```

```
/locus_tag="PF11  
/codon_start=1
```

```
/product="hypoth
/protein_id="AAN
```

/db\_xref="GI:234







QY 1182 CAAACGTAACGCAAAATATCTCTATTTTAAATTTTACACAAATATATAATCTT 1241  
 Db 56613 TAT 56555  
 QY 1242 CTATATTTAAATTTTACAAAT 1301  
 Db 56554 ATTAT 56495  
 QY 1302 AATTTTAACTTATGATATTTTATTTCTTAAATTTTGAATCTTCAATATATCTGATA 1361  
 Db 56494 AT 56435  
 QY 1362 TTTATTTTATATTTTGTGCAATTTTCTTATAGTTTGAATTTTACCTTTATCTTGCTG 1421  
 Db 56434 TTTAT 56375  
 QY 1422 AAACCTAATTTCAATATATAGATTTGTGAAGACACATTTGACATCTTGAACATTTGTT 1481  
 Db 56374 AAT 56315  
 QY 1482 TTAACCTTTGTGAATTTTAAAGTATATAAATTTGACATTTATGACCATTTATTAATA 1541  
 Db 56314 GTAT 56255  
 QY 1542 TACTTCCTTTGTCTTTTAAAAA 1564  
 Db 56254 TATATGTACACCTCTTGCAATA 56232

RESULT 12  
 AC131649 181976 bp DNA linear PRI 25-AUG-2002  
 LOCUS Homo sapiens chromosome 16 clone RP11-609N14, complete sequence.  
 AC131649 AC016139  
 DEFINITION AC131649.1 GI:22475305  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 TITLE 1 (bases 1 to 181976)  
 JOURNAL DOB Joint Genome Institute.  
 COMMENT Sequencing of Human Chromosome 16  
 Submitted (25-AUG-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 On Aug 25, 2002 this sequence version replaced GI:20279329.  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.  
 FEATURES  
 source Location/Qualifiers  
 1. 181976 /Organism="Homo sapiens"

BASE COUNT 51909 a 37229 c 38430 g 54408 t  
 ORIGIN  
 Query Match 5.1%; Score 90.2; DB 9; Length 181976;  
 Best Local Similarity 48.8%; Pred. No. 6e-05;  
 Matches 306; Conservative 0; Mismatches 313; Indels 8; Gaps 2;  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-609N14"

QY 912 ACTACCTTTAGTACATTTTATAGAAAATTTGATCAATCATTTATATATGCTCTTA 971  
 Db 174813 AT 174872  
 QY 972 TTTAAATTTATGATTAAGTTGTATCATTTAGATTTAGAAAACCAATATGCTCGCTTG 1031  
 Db 174873 TAT 174932  
 QY 1032 ATTTTGAATTTATGTTTTCATATGTTACTTTTCTCAAGCCTATATATAAATCTTGAT 1091  
 Db 174933 TAT 174992  
 QY 1092 GCTAATTTGTATGCTGGAAGAAAATGCTGAATTTGAATTTGAATTTATGATTTGCA 1151  
 Db 174993 TAT 175052  
 QY 1152 AGTCGCAATTCATGCAATGCAATTTAGTACAAAACGTACGCAAAATATCTCTAT 1211  
 Db 175053 AAT 175112  
 QY 1212 TTAATTTTACACAAATATATAAATTTCTGATTTTAAATTTTACAAATATATATAT 1271  
 Db 175113 AT 175172  
 QY 1272 ATCAGCTGTCACTTTAGAAATACCAACAATATTAATCTTATGATATTTATCTTAA 1331  
 Db 175173 ATAA----AT 175228  
 QY 1332 TATATTTAGAGTCTGCAATATATGATATTTATTTATTTTGTGTCATATTTCTTA 1391  
 Db 175229 TAAATATATTTAT 175284  
 QY 1392 TGTTTAGAGTTAACCTTATATCTGCTCAACAGTAAATTCATATATAGATTGCA 1451  
 Db 175285 TAT 175344  
 QY 1452 AGGACACATTTGACATCTTGAACAATTTGCTTTAACTTTGGAATGTTAAAGTAAATA 1511  
 Db 175345 TAT 175404  
 QY 1512 AACATTCAGATTTATGACCATCTTAA 1558  
 Db 175405 AAT 175431

RESULT 13  
 AC117444/c 104853 bp DNA linear PRI 31-JUN-2002  
 LOCUS Homo sapiens 3 BAC RP11-215L17 (Roswell Park Cancer Institute Human BAC Library) complete sequence.  
 AC117444  
 DEFINITION AC117444.6 GI:22024337  
 VERSION  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 REFERENCE  
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 TITLE 1 (bases 1 to 104853)  
 JOURNAL Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Albrooks,S.L., Amarantunge,H.C., Are,J.R., Ayale,M., Banks,T., Barbiana,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,



```

repeat_region      /rpt_family="L1M4"
19255_19775
/rpt_family="L1PA13"
repeat_region      /rpt_family="L1PA13"
20757_20815
/rpt_family="CP-rich"
repeat_region      complement(21111_27252)
/rpt_family="L1PA5"
22521_22798
/standard_name="D11S206"
27253_27294
/rpt_family="AT-rich"
repeat_region      complement(27534_28145)
/rpt_family="L1MC1"
28155_28203
/rpt_family="AT-rich"
repeat_region      complement(28585_28682)
/rpt_family="U6"
29288_29952
/rpt_family="MER76"
30941_30960
/rpt_family="(TTTC)n"
repeat_region      complement(30963_31250)
/rpt_family="A1Uy"
repeat_region      complement(31825_31911)
/rpt_family="I2"
repeat_region      complement(32569_32680)
/rpt_family="A1USq"
33002_33226
/rpt_family="L1R59"
repeat_region      /rpt_family="L1R59"
33247_33731
/rpt_family="MER72"
repeat_region      complement(33796_34214)
/rpt_family="I2"
repeat_region      complement(34233_35306)
/rpt_family="MER11C"
35706_35894
/rpt_family="MIR"
36208_36244
/rpt_family="AT-rich"

```

Query Match	5.0%;	Score 90;	DB 9;	Length 104853;
Best Local Similarity	45.4%;	Pred. No. 7.5e-05;		
Matches 399;	Conservative 0;	Mismatches 475;	Indels 4;	Gaps 2

QY	888	TAAATATATGCTGGTACACACAGTACCTTTGGTAGATTATAAGAAATTTGATC	947
Db	40503	TATATAATATTTATATATTTATATATAAATATATTTATATTTATATATAAATATAT	40444
QY	948	ATCATATTATATATTGTGCTTATTTPAAAATTATGATAAGTGTATCATTAAGATTGA	1007
Db	40443	TATATATTATATATATAAAATATATATATATATATATATATATAAATATATATATAT	40384
QY	1008	GAAATACCAATAGAGCTTCGCTGATTTTGGATATATGTCTTATGTTTCTTACTTTCTTC	1067
Db	40383	ATATATATAAAATATATATATATATATATATATATAATATATATATATATATAAATA	40324
QY	1068	AAGCCTATATATAAATCTTGTAAATGCTAAATGTATGCTGAAAATAATGTATAGAT	1127
Db	40323	TATATATATATATATATATATAAATATATATATATATATATATATATATATATATAT	40264
QY	1128	TCAATAGAAATATATGTAATTTCAAGTCGCAAAATCATCAATAGAAATTTAGTCGCAAAAC	1187
Db	40263	TATATATATAATTTAT	40204
QY	1188	GTAATCGAAAAATTTCTCTTATTTTAAATTTTACAATATATATAATTTCTCTTAT	1247
Db	40203	TTATATATATATATATATATATACATTTTATATATATATATATATATATATATATATA	40144
QY	1248	TTAAATTTTACAATATATATATATATATACCGTCGACCTTATAGAAATCCACCAATATAT	1306
Db	40143	TATATAATATATATAAATATATCTATATAAATATATATTTTATATAAATATATATATAT	40084
QY	1307	TAAATCTTAGATATTTTATTTCTTAAATAATTTTGAGATCTCTCAATATATCTGATTTTAT	1366

[illegible]

RESULT	14
Locus	BX248102/c
Definition	Danio rerio clone CH211-243F2, *** SEQUENCING IN PROGRESS ***, 2 unworded pieces.
Accession	BX248102
Version	BX248102.5 GI:29538864
Keywords	Htg; Htgs_Phasel; Htgs_DRAFT; Htgs_Fulltop. Danio rerio ( <i>zebrafish</i> ) Danio rerio
Organism Source	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprindae; Danio. 1 (bases 1 to 150236) McLaren S.
Reference Authors	Direct Submission
Title	Submitted (03-APR-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: <a href="mailto:zfish-help@sanger.ac.uk">zfish-help@sanger.ac.uk</a> Clone requests: <a href="mailto:clonerequests@sanger.ac.uk">clonerequests@sanger.ac.uk</a> On Apr 3, 2003 this sequence version replaced gi:28301010.
JOURNAL	
Comment	

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 130320: contig of 130320 bp in length  
 \* 130321 130420: gap of 100 bp  
 \* 130421 150236: contig of 19816 bp in length.

## FEATURES

## source

1. 150236  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="CH211-243F3"  
 /clone\_1lb="CHORI-211"  
 1. 130320  
 /note="assembly\_fragment:01135  
 fragment\_chain:1  
 clone\_end:SP6  
 vector\_side:left"  
 130421..150236  
 /note="assembly\_fragment:00144  
 fragment\_chain:1  
 clone\_end:T7  
 vector\_side:right"

## misc\_feature

1. 130320  
 /note="assembly\_fragment:00144  
 fragment\_chain:1  
 clone\_end:T7  
 vector\_side:right"

BASE COUNT 47331 a 26020 c 27386 g 49399 t 100 others  
 ORIGIN

## Query Match

5.0%; Score 90; DB 2; Length 150236;  
 Best Local Similarity 47.0%; Pred. No. 6.8e-05;  
 Matches 314; Conservative 0; Mismatches 350; Indels 4; Gaps 1;

Qy 786 GAGCACTTACAGACCTTTTACTTAAATCTACAAAGAGAGATTTTACCACTTAGA 845  
 Db 131133 GTGCAACATATGCGCTTTTAAAGATTAATTAACAACTGTATCTTAATATAGCTGT 131074  
 Qy 846 GAAGTAATGGAGTTTAAAGACACATTAAGGGAGCTTAAATTAATGCTGTA 905  
 Db 131073 ATATGCCACCAAGTTATAGTGAAGTCTTAACCTGCTGTTCCCAACATTTT 131014  
 Qy 906 ACCACACACTACCTTTAGTATGATTTATAGAAAATGTAATCATCATTAATTAATG 965  
 Db 131013 GCACCACTTCTAGTCTTATATATATATATATATATATATATATATATATAT 130954  
 Qy 966 TCCTTATTAATTAATGATTAAGTGTATCATTAAGATTGAGAAAACCAATAGCC 1025  
 Db 130953 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 130894  
 Qy 1026 GCTTGATTTTGAATATGTTTCTATGTTACTTTCTCAAGCCTATATAAAACT 1085  
 Db 130893 TTATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 130834  
 Qy 1086 TGTAAAGCTAAATGTATGCTGAAAAAATGTTAATGAATCAATAGAAATTAATG 1145  
 Db 130833 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 130774  
 Qy 1146 TTTTAAAGTCCAAATGCATCAATAGAAATTTAGTCAAAAACGTAACCTCAAAATTT 1205  
 Db 130773 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTAT 130714  
 Qy 1206 CTATTTTAAATTTTACAAATATA---AAAAATCTCTTAATTTAAATTTTCAAT 1261  
 Db 130713 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 130654  
 Qy 1262 AATTAATTTATCCCTGCTCCCTTTGAATACCAACAATTAATTAATTAATTAAT 1321  
 Db 130653 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 130554  
 Qy 1322 TTATCTTAATTAATTTGAGATCTCTCAATATATGATATTTATTTATTTATTTGCTCA 1381  
 Db 130553 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 130534  
 Qy 1382 TATTTCTATGTTTAAAGTTAAACCTTAATATCTTGCTCAACTAGTAATTAATTAAT 1441

Db 130533 TTATTTATTAATTAATTAATTTTATTAATTAATTAATTAATTAATTAATTAAT 130474  
 Qy 1442 GAGTTTGG 1449  
 Db 130473 TATTAATTAAT 130466

## RESULT 15

## LOCUS

BX119991 153326 bp DNA linear HTG 03-APR-2003  
 Danio rerio clone CH211-240A8, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 5  
 unorderd pieces.

## DEFINITION

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

BX119991 153326 bp DNA linear HTG 03-APR-2003  
 Danio rerio clone CH211-240A8, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 5  
 unorderd pieces.  
 BX119991  
 BX119991.2 GI:29538704  
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS\_FULLTOP.  
 Danio rerio (zebrafish)  
 Danio rerio  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;  
 Cypriniformes; Cyprinidae; Danio.  
 1 (bases 1 to 153326)  
 McLaren, S.  
 Direct Submission  
 Submitted (03-APR-2003) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk  
 On Apr 3, 2003 this sequence version replaced gi:27903718.  
 ----- Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: zfish-help@sanger.ac.uk  
 Project Information  
 Center project name: zc240A8  
 ----- Summary Statistics  
 Assembly program: XGAP4; Version 4.5  
 Chemistry: Dye-terminator; 100% of reads  
 Consensus quality: 151746 bases at least Q40  
 Consensus quality: 152254 bases at least Q30  
 Consensus quality: 152541 bases at least Q20  
 Insert size: 152926; sum-of-contigs  
 Insert size: 162217; 2.3% error; agarose-fp  
 Quality coverage: 6.73x in Q20 bases; sum-of-contigs Quality  
 coverage: 6.78x in Q20 bases; agarose-fp

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 5 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1 19915: contig of 19915 bp in length  
 \* 19916 20015: gap of 100 bp  
 \* 20016 25492: contig of 5477 bp in length  
 \* 25493 25592: gap of 100 bp  
 \* 25593 36617: contig of 11025 bp in length  
 \* 36618 36718: gap of 100 bp  
 \* 36718 135889: contig of 99272 bp in length  
 \* 135890 136089: gap of 100 bp  
 \* 136090 153326: contig of 17237 bp in length.

## FEATURES

## source

1. 153326  
 /organism="Danio rerio"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:7955"  
 /clone="CH211-240A8"  
 /clone\_1lb="CHORI-211"  
 1. 19915  
 /note="assembly\_fragment:01135  
 clone\_end:T7

## misc\_feature

```
misc_feature      vector size:left"
20016..25492
/Note="assembly_fragment:00605
fragment_chain:1"
misc_feature      23593..36617
/Note="assembly_fragment:01486
fragment_chain:1"
misc_feature      36718..135989
/Note="assembly_fragment:01039
fragment_chain:1"
misc_feature      136090..153326
/Note="assembly_fragment:01628
fragment_chain:1
clone_end:SP6
vector size:right"
BASE COUNT      50335 a 27839 c 26492 g 48257 t 403 others
ORIGIN
```

```
Query Match      5.0%; Score 90; DB 2; Length 153326;
Best Local Similarity 47.0%; Pred. No. 6,7e-05;
Matches 314; Conservative 0; Mismatches 350; Indels 4; Gaps 1;
```

```
QY      786 GAGCACTTACAGCACTTTTACCAAACTCAAGAGAGATTTTACCACTTGA 845
DB      19104 GTGCACATATGCGCTTTTAAGTAAATAACAACCTTGTACTTAATATAGGCTGT 19163
QY      846 GAAATATGGAGTTAAAGCAACATTAAGGGAGTGTAAATTAATGTGTGA 905
DB      19164 ATTATGCCACAGTTAATGAGACTGTATACCTTAACGTGTACCAACTTTTGA 19223
QY      906 ACCACCACTACCTTTAGTATATTAAGAAATGTATCATCATTAATATTTG 965
DB      19224 GCAGCCATTACTAGTCTATTAATTAATTAATTAATTAATTAATTAATTAAT 19283
QY      966 TCCTTATTTAAATTAATGAATAAGTGTATCATTAAGTGAAGAAACCAATATGCTGC 1025
DB      19284 ATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19343
QY      1026 GTCTTGAATTTTGAATTAATGTTTCTAGTTACTTTTCTCAAGCCATTAATAAACTT 1085
DB      19344 TTATTTTAATTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 19403
QY      1086 TGTATGCTAAATGTGCTGGAAGAAATGCTAATGAATCAATAGAAATTAATGTA 1145
DB      19404 TATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 19463
QY      1146 TTCAAAGTCAAAATCCATCAATAGAAATTAATTAATTAATTAATTAATTAATTC 1205
DB      19464 TATTAATTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 19523
QY      1206 CTATTTTAATTTTACAAATATA---AAATATCTCTATTTTAATTTTACAT 1261
DB      19524 ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19583
QY      1262 AATAATTAATTAACCTGTCACTTAGAATACCAACAATTAATTAATTAATTAATTA 1321
DB      19584 ATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19643
QY      1322 TTATCTTAATTAATTTGAGATCTTCAATATATGATATTTATTTATATTTGTCGA 1381
DB      19644 TTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 19703
QY      1382 TATTTCTTAATTTAGATTAACTTATATCTTGTCGAACCTAGTAAATCAATATAT 1441
DB      19704 TTATTTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 19763
QY      1442 GAGTTTGT 1449
DB      19764 TATATTTAT 19771
```

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 16:23:54 : Search time: 2616 Seconds

(Without alignments)  
16565.321 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atccacctctgaccttga.....tagagcgatcgaagcgaacc 1783

Scoring table: OLIGO NUC

Gapop 60.0, Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST.\*  
1: em\_estda.\*  
2: em\_estlum.\*  
3: em\_estlin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_hnc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hnc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estlum.\*  
16: em\_estcom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_hiv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_mam.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rtd.\*  
26: em\_gss\_phg.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*  
29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	1.3	625	28	AZ978994
2	23	1.3	189	10	BA422474
3	23	1.3	575	12	BU562777
4	23	1.3	600	12	BU563982

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
AZ978994	625 bp	DNA	linear	GSS	27-APR-2001							
2X0255104R	Mouse 10kb	plasmid	UUGC2M	library	Mus musculus genomic							
clone UUGC2M0255104 R					genomic survey sequence.							
AZ978994												
AZ978994.1	GI:13850221											
GSS:												
Mus musculus (house mouse)												
Mus musculus												
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.												
1 (bases 1 to 625)												
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamill, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A. and Wright, D., Weiss, R.												
Mouse whole genome scaffolding with paired end reads from 10kb plasmid inserts												
Unpublished												
Contact: Robert B. Weiss												
University of Utah Genome Center												
University of Utah												
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT 84112, USA												
Tel: 801 585 5606												
Fax: 801 585 7177												

## ALIGNMENTS

C	5	23	1.3	708	29	BZ2323523
C <td>6</td> <td>22</td> <td>1.2</td> <td>342</td> <td>28</td> <td>BH780793</td>	6	22	1.2	342	28	BH780793
C <td>7</td> <td>22</td> <td>1.2</td> <td>429</td> <td>14</td> <td>CA571794</td>	7	22	1.2	429	14	CA571794
C <td>8</td> <td>22</td> <td>1.2</td> <td>479</td> <td>13</td> <td>BQ562120</td>	8	22	1.2	479	13	BQ562120
C <td>9</td> <td>22</td> <td>1.2</td> <td>495</td> <td>28</td> <td>BH787917</td>	9	22	1.2	495	28	BH787917
C <td>10</td> <td>22</td> <td>1.2</td> <td>592</td> <td>10</td> <td>BF198774</td>	10	22	1.2	592	10	BF198774
C <td>11</td> <td>22</td> <td>1.2</td> <td>653</td> <td>29</td> <td>BZ421643</td>	11	22	1.2	653	29	BZ421643
C <td>12</td> <td>22</td> <td>1.2</td> <td>1136</td> <td>14</td> <td>CA995790</td>	12	22	1.2	1136	14	CA995790
C <td>13</td> <td>22</td> <td>1.2</td> <td>1136</td> <td>29</td> <td>CC283036</td>	13	22	1.2	1136	29	CC283036
C <td>14</td> <td>21</td> <td>1.2</td> <td>196</td> <td>9</td> <td>AU060103</td>	14	21	1.2	196	9	AU060103
C <td>15</td> <td>21</td> <td>1.2</td> <td>272</td> <td>13</td> <td>BQ851931</td>	15	21	1.2	272	13	BQ851931
C <td>16</td> <td>21</td> <td>1.2</td> <td>371</td> <td>28</td> <td>AZ845955</td>	16	21	1.2	371	28	AZ845955
C <td>17</td> <td>21</td> <td>1.2</td> <td>423</td> <td>10</td> <td>BG008892</td>	17	21	1.2	423	10	BG008892
C <td>18</td> <td>21</td> <td>1.2</td> <td>438</td> <td>28</td> <td>AZ003018</td>	18	21	1.2	438	28	AZ003018
C <td>19</td> <td>21</td> <td>1.2</td> <td>488</td> <td>9</td> <td>A1700637</td>	19	21	1.2	488	9	A1700637
C <td>20</td> <td>21</td> <td>1.2</td> <td>501</td> <td>12</td> <td>B1867863</td>	20	21	1.2	501	12	B1867863
C <td>21</td> <td>21</td> <td>1.2</td> <td>510</td> <td>10</td> <td>BF814093</td>	21	21	1.2	510	10	BF814093
C <td>22</td> <td>21</td> <td>1.2</td> <td>511</td> <td>12</td> <td>BM336511</td>	22	21	1.2	511	12	BM336511
C <td>23</td> <td>21</td> <td>1.2</td> <td>531</td> <td>14</td> <td>CB162751</td>	23	21	1.2	531	14	CB162751
C <td>24</td> <td>21</td> <td>1.2</td> <td>534</td> <td>9</td> <td>AL922682</td>	24	21	1.2	534	9	AL922682
C <td>25</td> <td>21</td> <td>1.2</td> <td>542</td> <td>29</td> <td>BZ781272</td>	25	21	1.2	542	29	BZ781272
C <td>26</td> <td>21</td> <td>1.2</td> <td>557</td> <td>13</td> <td>BQ851691</td>	26	21	1.2	557	13	BQ851691
C <td>27</td> <td>21</td> <td>1.2</td> <td>562</td> <td>29</td> <td>BZ779081</td>	27	21	1.2	562	29	BZ779081
C <td>28</td> <td>21</td> <td>1.2</td> <td>603</td> <td>10</td> <td>BF507982</td>	28	21	1.2	603	10	BF507982
C <td>29</td> <td>21</td> <td>1.2</td> <td>622</td> <td>13</td> <td>BQ850954</td>	29	21	1.2	622	13	BQ850954
C <td>30</td> <td>21</td> <td>1.2</td> <td>636</td> <td>29</td> <td>BX168297</td>	30	21	1.2	636	29	BX168297
C <td>31</td> <td>21</td> <td>1.2</td> <td>647</td> <td>12</td> <td>BJ371907</td>	31	21	1.2	647	12	BJ371907
C <td>32</td> <td>21</td> <td>1.2</td> <td>648</td> <td>12</td> <td>BJ340558</td>	32	21	1.2	648	12	BJ340558
C <td>33</td> <td>21</td> <td>1.2</td> <td>654</td> <td>29</td> <td>BZ869284</td>	33	21	1.2	654	29	BZ869284
C <td>34</td> <td>21</td> <td>1.2</td> <td>662</td> <td>12</td> <td>BG776697</td>	34	21	1.2	662	12	BG776697
C <td>35</td> <td>21</td> <td>1.2</td> <td>671</td> <td>28</td> <td>AZ516262</td>	35	21	1.2	671	28	AZ516262
C <td>36</td> <td>21</td> <td>1.2</td> <td>676</td> <td>14</td> <td>CA448878</td>	36	21	1.2	676	14	CA448878
C <td>37</td> <td>21</td> <td>1.2</td> <td>694</td> <td>13</td> <td>BU028797</td>	37	21	1.2	694	13	BU028797
C <td>38</td> <td>21</td> <td>1.2</td> <td>699</td> <td>9</td> <td>AU060686</td>	38	21	1.2	699	9	AU060686
C <td>39</td> <td>21</td> <td>1.2</td> <td>703</td> <td>29</td> <td>AG142042</td>	39	21	1.2	703	29	AG142042
C <td>40</td> <td>21</td> <td>1.2</td> <td>708</td> <td>13</td> <td>BU719850</td>	40	21	1.2	708	13	BU719850
C <td>41</td> <td>21</td> <td>1.2</td> <td>713</td> <td>12</td> <td>BJ377372</td>	41	21	1.2	713	12	BJ377372
C <td>42</td> <td>21</td> <td>1.2</td> <td>715</td> <td>12</td> <td>BJ386295</td>	42	21	1.2	715	12	BJ386295
C <td>43</td> <td>21</td> <td>1.2</td> <td>728</td> <td>13</td> <td>BQ988528</td>	43	21	1.2	728	13	BQ988528
C <td>44</td> <td>21</td> <td>1.2</td> <td>730</td> <td>13</td> <td>BQ995401</td>	44	21	1.2	730	13	BQ995401
C <td>45</td> <td>21</td> <td>1.2</td> <td>731</td> <td>12</td> <td>BT402629</td>	45	21	1.2	731	12	BT402629



Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Convolvulaceae; Ipomoea.

1 (bases 1 to 575)

Hoshino,A., Seki,M., Shin-I,T., Carninci,P., Kamiya,A., Shiraki,T., Nitaaka,E., Shinozaki,K., Hayashizaki,Y., Kohara,Y. and Iida,S.

ESTs of Japanese morning glory

Unpublished

Contact: Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@gens.nig.ac.jp.

Location/Qualifiers

1..575

/organism="Ipomoea nil"

/mol\_type="mRNA"

/cultivar="Tokyo-kokei standard"

/db\_xref="taxon:35883"

/clone="j1m37p16"

/tissue\_type="mixture of flower and flower bud"

/clone\_lib="Ipomoea nil mixture of flower and flower bud"

141 a 146 c 93 g 195 t

BASE COUNT

141 a 146 c 93 g 195 t

ORIGIN

Query Match 1.3%; Score 23; DB 12; Length 575;

Best Local Similarity 100.0%; Pred. No. 8.3;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1718 TCCGTCCTCTTTCTTTCTTCAC 1740

|||||

527 TCCGTCCTCTTTCTTTCTTCAC 549

Db

RESULT 4

LOCUS B1563982

DEFINITION B1563982 Ipomoea nil mixture of flower and flower bud Ipomoea nil

ACCESSION B1563982

VERSION B1563982.1

KEYWORDS

SOURCE

ORGANISM

EST.

Ipomoea nil (Japanese morning glory)

Ipomoea nil

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; lamids; Solanales; Convolvulaceae; Ipomoea.

1 (bases 1 to 600)

Hoshino,A., Seki,M., Shin-I,T., Carninci,P., Kamiya,A., Shiraki,T., Nitaaka,E., Shinozaki,K., Hayashizaki,Y., Kohara,Y. and Iida,S.

ESTs of Japanese morning glory

Unpublished

Contact: Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshin@gens.nig.ac.jp.

Location/Qualifiers

1..600

/organism="Ipomoea nil"

/mol\_type="mRNA"

/cultivar="Tokyo-kokei standard"

/db\_xref="taxon:35883"

/clone="j1m40n16"

/tissue\_type="mixture of flower and flower bud"

/clone\_lib="Ipomoea nil mixture of flower and flower bud"

150 a 151 c 96 g 200 t

BASE COUNT

150 a 151 c 96 g 200 t

ORIGIN

Query Match 1.3%; Score 23; DB 12; Length 600;

Best Local Similarity 100.0%; Pred. No. 8.4;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1718 TCCGTCCTCTTTCTTTCTTCAC 1740

|||||

527 TCCGTCCTCTTTCTTTCTTCAC 549

Db

RESULT 5

LOCUS B2323523/c

DEFINITION B2323523 708 bp DNA linear GSS 06-NOV-2002

1a79c09.g1 WGS-Zmaysf (JM107 adapted methyl filtered) Zea mays

genomic clone 1a79c09 5', genomic survey sequence.

ACCESSION B2323523

VERSION B2323523.1

KEYWORDS

SOURCE

ORGANISM

Zea mays

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 708)

Rabinowitz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N., Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L., Zuttervern,T., McCombie,W.R. and Martienssen,R.A.

Genomic shotgun sequences from Zea mays (methyl-filtered)

Unpublished

Contact: W. Richard McCombie

Lita Annenberg Hazen Genome Sequencing Center

Cold Spring Harbor Laboratory

PO Box 100, Cold Spring Harbor, NY 11724, USA

Tel: 516 367 8884

Fax: 516 367 8874

Email: mcombie@cshl.org

Plate: 1a79 row: c column: 09

Seq primer: -21M13UnivRev

Class: shotgun

High quality sequence stop: 708.

Location/Qualifiers

1..708

/organism="Zea mays"

/mol\_type="genomic DNA"

/cultivar="B73"

/db\_xref="taxon:4577"

/clone="1a79c09"

/lab\_host="JM107 or DH5a"

/clone\_lib="WGS-Zmaysf (JM107 adapted methyl filtered)"

/note="Organ: immature ears; Site\_1: Xba I; Site\_2: Xba I; The vector was digested with Xba I and one nucleotide was added by fill in in the recessive 3' end. The genomic DNA was nebulized, end repaired, adaptor ligated and size fractionated using sephadex. The resulting fragments were between 0.8 and 3 kb and were cloned into the vector (x/y reads in M13mp19, b/g reads in pUC19). The same ligation was transformed in either JM107 or DH5a."

172 a 178 c 139 g 219 t

BASE COUNT

172 a 178 c 139 g 219 t

ORIGIN

Query Match 1.3%; Score 23; DB 29; Length 708;

Best Local Similarity 100.0%; Pred. No. 8.6;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1211 TTAAATTTCACACATATATA 1233

|||||

506 TTAAATTTCACACATATATA 484

Db

RESULT 6

LOCUS B1780793

DEFINITION B1780793 342 bp DNA linear GSS 28-MAR-2002

fzmb020f010g10f0 famb filtered library Zea mays genomic clone

fzmb020f010g10 5', genomic survey sequence.

ACCESSION BH780793  
 VERSION BH780793.1 GI:19783834  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays

REFERENCE Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Zea.  
 1 (bases 1 to 342)  
 AUTHORS Budiman, M.A., Freese, R.G., Bedell, J.A., Nunberg, A.N. and Lakey, N.D.  
 TITLE Geneshresher methylation filtered genomic sequences from maize  
 JOURNAL Unpublished  
 COMMENT Contact: Bedell JA  
 Orion Genomics, LLC  
 4041 Forest Park Ave, St. Louis, MO 63108, USA  
 Tel: 314 615 6979  
 Fax: 314 615 5975  
 Email: jbedell@oriongenomics.com  
 Plate: fmb020f010 row: g column: 10  
 Seq primer: M13 forward  
 Class: shotgun  
 High quality sequence stop: 342.

FEATURES  
 source  
 1..342  
 location/Qualifiers  
 /organism="Zea mays"  
 /mol\_type="Genomic DNA"  
 /cultivar="M017"  
 /db\_xref="taxon:4577"  
 /clone\_lib="fmb020f010g10"  
 /note="Organ: leaf; Vector: pGSK(-); Site: 1; HincII; DNA prepared from purified nuclei was randomly sheared, end-repaired, size fractionated to enrich for the 0.5 to 5 kb fraction, ligated into HincII-digested pGSK(-) vector and electroporated into E. coli cells."

BASE COUNT 111 a 59 c 69 g 103 t  
 ORIGIN

Query Match 1.2%; Score 22; DB 28; Length 342;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 750 GAAGGAATGTGCACTT 771  
 118 GAAGGAATGTGCACTT 139

RESULT 7  
 LOCUS CA571794 429 bp mRNA linear EST 19-NOV-2002  
 DEFINITION K0524H02-SN NIA Mouse Hematopoietic Stem Cell (Lin-/C-Kit+/Sca-1+)  
 CDNA Library (Long) Mus musculus CDNA clone NIA:K0524H02  
 IMAGE:30066229 5', mRNA sequence.  
 CA571794  
 CA571794.1 GI:25116495

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 EST  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 429)  
 AUTHORS Piao, Y., Kargul, G.J., Dudekula, D.B., Qian, Y., Luo, A., Carter, M.G., Taub, D., Longo, D.L., Keller, J. and Ko, M.S.H.  
 TITLE Systematic Analyses of NIA Mouse Hematopoietic Stem Cell (Lin-/C-Kit+/Sca-1+) CDNA Library (Long)  
 JOURNAL Unpublished  
 Other\_ESTS: K0524H02-3  
 Contact: Dawood B. Dudekula  
 Laboratory of Genetics  
 National Institute on Aging/National Institutes of Health  
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA  
 Email: cdna@igsun.grc.nia.nih.gov  
 Plate: K0524 row: H column: 02

Seq primer: M13 Reverse  
 High quality sequence stop: 429  
 POLYA=No.  
 FEATURES  
 source  
 1..429  
 location/Qualifiers  
 /organism="Mus musculus"  
 /mol\_type="mRNA"  
 /strain="C57BL/6NCr"  
 /db\_xref="taxon:10090"  
 /clone="NIA:K0524H02 IMAGE:30066229"  
 /tissue\_type="Hematopoietic Stem Cell (Lin-/C-Kit+/Sca-1+)"  
 /dev\_stage="Age approx. 10 weeks old"  
 /lab\_host="DH10B"  
 /clone\_lib="NIA Mouse Hematopoietic Stem Cell (Lin-/C-Kit+/Sca-1+) CDNA Library (Long)"  
 /note="Vector: pSPORT1 (Invitrogen); Site: 1; SalI; Site: 2; NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.grc.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). PMID: 11544199). Total RNAs were obtained from Drs. Dennis Taub, Dan Longo (National Institute on Aging, USA), Jonathan Keller (National Cancer Institute, USA). Double-stranded cDNAs were synthesized with an Oligo(dT) primer (Invitrogen): 5'-pGACTAGTCTAGATCGGACGCGCCCTTTTCTTTT-3' from 4.8 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to lops-linker IL-SalI, purified by phenol/chloroform, and separated from free linkers by centrifugation. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer SalI-S. The products were purified by phenol/chloroform and centrifugation. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DH10B E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.7 kb. The library was constructed by Yulan Piao (NIA)."

BASE COUNT 99 a 102 c 88 g 140 t  
 ORIGIN

Query Match 1.2%; Score 22; DB 14; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1560 AAAAAGTGTGCAAAATGCT 1581  
 61 AAAAAGTGTGCAAAATGCT 40

RESULT 8  
 LOCUS B0562120 479 bp mRNA linear EST 20-JUN-2002  
 DEFINITION H074C02-5 NIA Mouse 7.4K cDNA Clone Set Mus musculus cDNA clone  
 H4074C02 5', mRNA sequence.  
 B0562120  
 B0562120.1 GI:21463006

ACCESSION  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 EST  
 Mus musculus (house mouse)  
 Mus musculus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 479)  
 AUTHORS VanBuren, V., Piao, Y., Dudekula, D.B., Qian, Y., Carter, M.G., Martin, P.R., Stagg, C.A., Baesey, U., Alba, K., Hamatani, T., Kargul, G.J., Luo, A.G., Kelso, J., Hide, W. and Ko, M.S.H.  
 TITLE Assembly, verification, and initial annotation of NIA 7.4K mouse cDNA clone set  
 JOURNAL Genome Res. 12 (12), 1999-2003 (2002)  
 MEDLINE 22354164

PUBMED  
12466305  
Other ESTs: H4074C02-3  
Contact: Yong Qian  
Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA  
Email: cdna@gsun.grc.nia.nih.gov  
This clone set has been freely distributed to the community. Please  
visit [http://lgsun.grc.nia.nih.gov/cdna/NIA\\_7\\_4k.html](http://lgsun.grc.nia.nih.gov/cdna/NIA_7_4k.html) for details.  
Plate: H4074 row: C column: 02  
Seq primer: -21M13 Reverse  
High quality sequence stop: 479  
POLYA=No.

FEATURES  
source  
Location/Qualifiers  
1..479  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6"  
/db\_xref="taxon:10090"  
/db\_xref="E01090"  
/clone="H4074C02"  
/sex="mixed"  
/dev\_stage="mixed"  
/lab\_host="DH10B"  
/clone\_lib="NIA Mouse 7.4k cDNA Clone Set"  
/note="Vector: pSPORT1; Site 1: SalI; Site 2: NotI; This  
clone is among a rearranged set of 7,407 clones from more  
than 20 cDNA libraries."

BASE COUNT  
110 a 117 c 95 g 157 t

Query Match 1.2%; Score 22; DB 13; Length 479;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 1560 AAAAAGTGTGATGAAATGCT 1581  
61 AAAAAGTGTGATGAAATGCT 40

RESULT 9  
BH787917/c 495 bp DNA linear GSS 28-MAR-2002  
LOCUS fmb020f010g10k0 fmb filtered library Zea mays genomic clone  
DEFINITION fmb020f010g10 5', genomic survey sequence.  
BH787917  
ACCSSION BH787917.1 GI:19795777  
VERSION GSS.  
KEYWORDS Zea mays  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoidae; Andropogoneae; Zea.  
1 (bases 1 to 495)  
Bushman, M.A., Freese, R.G., Bedell, J.A., Nunberg, A.N. and Lakey, N.D.  
GenetResher methylation filtered genomic sequences from maize  
Unpublished  
Contact: Bedell JA  
Otion Genomics, LLC  
4041 Forest Park Ave, St. Louis, MO 63108, USA  
Tel: 314 615 6979  
Fax: 314 615 5975  
Email: jbedell@otiongenomics.com  
Plate: fmb020f010 row: 5 column: 10  
Seq primer: SK reverse  
Class: shotgun  
High quality sequence stop: 495.

FEATURES  
source  
Location/Qualifiers  
1..495  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/cultivar="MO1.7"  
/db\_xref="taxon:4577"

BASE COUNT  
154 a 102 c 79 g 160 t

Query Match 1.2%; Score 22; DB 26; Length 495;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB 750 GAAGGAATGTGAGTACCTT 771  
404 GAAGGAATGTGAGTACCTT 383

RESULT 10  
BF198774 592 bp mRNA linear EST 03-NOV-2000  
LOCUS 248732 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.  
DEFINITION BF198774  
ACCESSION BF198774.1 GI:11090224  
VERSION EST.  
KEYWORDS Sus scrofa (pig)  
SOURCE Sus scrofa  
ORGANISM Sus scrofa

REFERENCE  
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
1 (bases 1 to 592)  
Fahrenkrug, S.C., Smith, T.P.L., Freking, B.A., Cho, J., White, J.,  
Vallet, J., Wise, J., Rohrer, G.A., Petrea, G., Sultana, R., Quackenbush  
J. and Keeler, J.W.  
Porcine gene discovery by normalized cDNA-library sequencing and  
EST cluster assembly  
Mamm. Genome 13 (8), 475-478 (2002)  
JOURNAL 22233789  
MEDLINE 1226715  
PUBMED 1226715  
COMMENT Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390  
Email: smith@email.marc.usda.gov  
Single pass sequencing. Bases called and alt trimmed with phred  
v0.980904.e. Vector identified by cross\_match with the -minscore 18  
and -mismatch 12 options.  
PCR Primers  
FORWARD: AGGAACAGCTATGACCAT  
BACKWARD: GTTCCACGACGACG  
Plate: 80 row: 1 column: 5  
Seq primer: ATTAGGTGACATATAG.

FEATURES  
source  
Location/Qualifiers  
1..592  
/organism="Sus scrofa"  
/mol\_type="mRNA"  
/db\_xref="taxon:9823"  
/clone\_lib="DH10B"  
/lab\_host="DH10B"  
/clone\_lib="MARC 2P1G"  
/note="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;  
library made from pooled tissue from testis, ovary,  
endometrium, hypothalamus, pituitary, and placenta."

BASE COUNT  
179 a 88 c 130 g 195 t

Query Match 1.2%; Score 22; DB 10; Length 592;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1384 TTTTCTATGTTTACGTTAA 1405  
TTTTTCTATGTTTACGTTAA 1405

Db 39 TTTCTATGTTTAGAGTTAA 60

RESULT 11  
B2421643  
LOCUS

DEFINITION B2421643 653 bp DNA linear GSS 10-DEC-2002  
h232f07.b1 WGS-bicolor (DH5a methyl filtered) Sorghum bicolor  
Genomic clone h232f07 5', genomic survey sequence.

ACCESSION B2421643  
VERSION B2421643.1 GI:26369150

KEYWORDS GSS.  
SOURCE Sorghum bicolor (sorghum)

ORGANISM Sorghum bicolor  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Sorghum.

REFERENCE  
AUTHORS

1 (bases 1 to 653)  
Rabinowicz, P.D., O'Shaughnessy, A.L., Ballia, V., Dedina, N.,  
Katzemburger, F., King, L., Miller, B., Nascimben, L.,  
Zutavern, T., Palmer, L., McCombie, W.R. and Martienssen, R.A.  
Genomic shotgun sequences from Sorghum bicolor (methyl-filtered)  
Unpublished

TITLE JOURNAL  
COMMENT Contact: W. Richard McCombie  
Lita Annenberg Hazen Genome Sequencing Center  
Cold Spring Harbor Laboratory  
PO Box 100, Cold Spring Harbor, NY 11724, USA  
Tel: 516 367 8884  
Fax: 516 367 8874  
Email: mcombie@cshl.org  
Plate: h232 row: f column: 07  
Seq primer: -21M13UnivFwd  
Class: shotgun  
High quality sequence stop: 653.  
Location/Qualifiers

FEATURES  
source

1..653  
/organism="Sorghum bicolor"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:4558"  
/clone="h232f07"  
/lab\_host="DH5a"  
/note="Site 1: Xba I; Site 2: Xba I; The vector was  
digested with XbaI and one nucleotide was added by fill in  
in the recessive 3' end. The genomic DNA was nebulized,  
end repaired, adaptor ligated and size fractionated using  
sephadex. The resulting fragments were between 0.8 and 3  
kb and were cloned into the vector (x/y reads in M13mp19,  
.b/g reads in pUC19). The same ligation was transformed  
into DH5a."  
BASE COUNT 190 a 135 c 121 g 206 t 1 others  
ORIGIN

Query Match 1.2%; Score 22; DB 29; Length 653;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 357 AGTCTGAACCTGTCATATTC 378  
Db 289 AGTTCTGAACCTGTCATATTC 310

RESULT 12  
CA995790  
LOCUS

DEFINITION CA995790 661 bp mRNA linear EST 07-JAN-2003  
r96a11.v1 Meloidogyne hapla c2 pAMP1 v1 Meloidogyne hapla cDNA 5',  
mRNA sequence.

ACCESSION CA995790  
VERSION CA995790.1 GI:27540661

KEYWORDS EST.  
SOURCE Meloidogyne hapla  
ORGANISM Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;  
Tylenchoidea; Heterodermidae; Meloidogyninae; Meloidogyne.

REFERENCE  
AUTHORS

1 (bases 1 to 661)  
McCarter, J., Clifton, S., Chiappelli, B., Pape, D., Martin, J., Wylie, T.,  
Dante, M., Maira, M., Hillier, L., Kuabla, T., Theising, B., Bowers, Y.,  
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, R.,  
Romko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,  
M., Allen, M., Pearson, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,  
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterson, R. and  
Wilson, R.  
The Washington Univ. Nematode EST Project, 1999  
Unpublished

TITLE JOURNAL  
COMMENT

Contact: McCarter JP  
The Washington Univ. Nematode EST Project, 1999  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
The library was constructed by Claire Murphy and Dr. James McCarter  
at Washington University, St. Louis. J2 were provided by Dr.  
Valerie Williamson of the University of California at Davis  
(vwilliamson@ucdavis.edu).  
Seq primer: -40RP from Gibco  
High quality sequence stop: 418.  
Location/Qualifiers

FEATURES  
source

1..661  
/organism="Meloidogyne hapla"  
/mol\_type="mRNA"  
/db\_xref="taxon:6305"  
/dev\_stage="J2"  
/lab\_host="DH10B"  
/clone="pAMP1 (Gibco)"  
/note="Vector: pAMP1 (Gibco); Site 1: NotI; Site 2: SalI;  
The library was constructed by Claire Murphy and Dr. James  
McCarter at Washington University, St. Louis. The cDNA was  
made by using Dynabead Oligo-dT priming (Dyna), PCR based  
library using a modified protocol from the SMART PCR cDNA  
Synthesis Kit from Clontech. Directionally cloned into the  
UDG sites of pAMP1. J2 were provided by Dr. Valerie  
Williamson of the University of California at Davis  
(vwilliamson@ucdavis.edu)."  
BASE COUNT 232 a 75 c 67 g 287 t

ORIGIN  
Query Match 1.2%; Score 22; DB 14; Length 661;  
Best Local Similarity 100.0%; Pred. No. 27;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1243 TTATTTTAAATTTTACATTAAT 1264  
Db 504 TTATTTTAAATTTTACATTAAT 525

RESULT 13  
CC283036/c  
LOCUS

DEFINITION CC283036 1136 bp DNA linear GSS 13-MAY-2003  
CH261-112N11.5p6.1 CH261 Gallus gallus genomic clone CH261-112N11,  
genomic survey sequence.

ACCESSION CC283036  
VERSION CC283036.1 GI:30649133

KEYWORDS GSS.  
SOURCE Gallus gallus (chicken)

ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
Phasianinae; Gallus.  
1 (bases 1 to 1136)  
Kremetzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,  
Warren, W., Graves, T., Wards, E. and Wilson, R.

REFERENCE  
AUTHORS  
TITLE JOURNAL  
COMMENT

Unpublished  
Contact: Richard K. Wilson  
Genome Sequencing Center  
Washington University School of Medicine

Email: submissions@watson.wustl.edu  
Insert Length: 18200 Std Error: 0.00  
Seq primer: SP6 ATTAGGTGACACTATG  
Class: BAC ends  
High quality sequence start: 73  
High quality sequence stop: 682.

## FEATURES

## source

1. 1136  
/organism="Gallus gallus"

/mol\_type="genomic DNA"

/strain="Red Jungle Fowl"

/db\_xref="taxon:9031"

/clone="CH261-112N11"

/sex="female"

/cell\_line="UCD001, indred 256"

/clone\_lib="CH261"

/note="Vector: pPARAC2.1; Site 1: EcoRI; Site 2: EcoRI; CH261 Female Chicken library - for library and clone ordering information: <http://www.chori.org/bacpac>"

BASE COUNT 338 a 208 c 174 g 415 t 1 others

## ORIGIN

Query Match 1.2%; Score 22; DB 29; Length 1136;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 388 AGAAAAATTTCTAGTGTGTTAG 409

DB 894 AGAAAAATTTCTAGTGTGTTAG 873

RESULT 14  
AU060103/c 196 bp mRNA linear EST 20-MAY-1999

LOCUS AU060103 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium

DEFINITION AU060103 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium

ACCESSION AU060103 GI:4881207

VERSION AU060103.1 GI:4881207

KEYWORDS Dictyostelium discoideum

SOURCE Dictyostelium discoideum

ORGANISM Dictyostelium discoideum

REFERENCE 1 (bases 1 to 196)

AUTHORS Morio,T., Urushihara,H., Satou,T., Ugawa,Y., Mizuno,H., Yoshida,M.,

Yoshino,R., Mitra,B.N., Pi,M., Satou,T., Takeuchi,K., Yasukawa,H.,

Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.

Developmental cDNA in Dictyostelium discoideum

Unpublished

Contact: Hideko Urushihara

Institute of Biological Sciences

University of Tsukuba

1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8572, Japan

Tel: 81-298-53-4664

Fax: 81-298-53-6614

Email: hideko@biol.tsukuba.ac.jp

PROJECT = Dictyostelium discoideum cDNA project in Japan.

Location/Qualifiers

1. 196

/organism="Dictyostelium discoideum"

/mol\_type="mRNA"

/strain="XX4"

/db\_xref="taxon:44689"

/clone="SLA320"

/dev\_stage="slug"

/clone\_lib="Dictyostelium discoideum SL (H.Urushihara)"

BASE COUNT 98 a 36 c 15 g 47 t

## ORIGIN

Query Match 1.2%; Score 21; DB 9; Length 196;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1032 ATTTTGAAATTAATGTTTCT 1052

DB 66 ATTTTGAAATTAATGTTTCT 46

RESULT 15  
B0851931/c 272 bp mRNA linear EST 14-AUG-2002

LOCUS B0851931/c OGB16K21, mRNA sequence.

DEFINITION OGB16K21, mRNA sequence.

ACCESSION B0851931 GI:22237396

VERSION B0851931.1 GI:22237396

KEYWORDS EST.

SOURCE Lactuca sativa

ORGANISM Lactuca sativa

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

Asteridae; Campanulidae; Asterales; Asteraceae; Cichorioideae;

Cichorieae; Lactuca.

1 (bases 1 to 272)

Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Riesberg,L.,

Lin,H., Van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison

,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z.,

Church,S., Jackson,L. and Bradford,K.

Lettuce and Sunflower ESTs from the Composite Genome Project

<http://comgenomics.ucdavis.edu/>

Unpublished

Contact: Alexander Kozik [R.W.Michelmore]

Department of Vegetable Crops, R.W.Michelmore Lab

University of California at Davis (UCD)

Asmunden Hall, UCD, Davis, CA 95616, USA

Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659

Email: akozik@ucdavis.edu [michelmore@vegmail.ucdavis.edu]

belongs to contig OG\_CA\_Contig4823, see <http://cgdb.ucdavis.edu/>

for details.

Plate: OGB16 row: K column: 21.

Location/Qualifiers

1. 272

/organism="Lactuca sativa"

/mol\_type="mRNA"

/cultivar="Salinas"

/db\_xref="taxon:4236"

/clone="OGB16K21"

/lab\_host="E.coli"

/clone\_lib="OG ABCDI lettuce salinas"

/note="Vector: pBRCDNA51AB. The library was constructed

from 10 different sources of RNA from a single genotype.

Separate cDNAs were generated using primers that

incorporated unique 5' and 3' tags to distinguish each

source of RNA. cDNAs were then pooled, size-fractionated,

directionally cloned into a custom medium-copy vector and

transformations made with four size classes to minimize

size bias. Details of each source of RNA and library

construction can be obtained at <http://cgdb.ucdavis.edu/>

TAG LIB=OG ABCDI lettuce salinas

TAG TTSSU=flowers post-fertilized

BASE COUNT 96 a 66 c 38 g 72 t

## ORIGIN

Query Match 1.2%; Score 21; DB 13; Length 272;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 493 TGATCTTGAAATTTGATGTT 513

DB 165 TGATCTTGAAATTTGATGTT 145

Search completed: November 25, 2003, 19:00:34  
Job time : 2621 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 18:16:43 ; Search time 440 Seconds

(without alignments)  
13350.983 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atcaacttcgacgttga.....tagagcagcagcgaac 1783

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 219069 seqs, 1647345023 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4380138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/2/pubpna/PCR\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/2/pubpna/PCRUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/2/pubpna/US09\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/2/pubpna/US10\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:\*  
16: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1.2	1067	12	US-10-027-632-250060	Sequence 250060.
2	1.2	1067	13	US-10-027-632-250060	Sequence 250060.
3	1.2	7400	12	US-09-960-858-4	Sequence 4, Appli
4	1.2	7400	12	US-09-960-870-4	Sequence 4, Appli
5	1.2	8011	12	US-10-311-455-52	Sequence 52, Appli
6	1.2	580073	12	US-10-027-632-220-1	Sequence 1, Appli
7	1.1	605	12	US-10-027-632-210764	Sequence 210764.
8	1.1	605	13	US-10-027-632-210764	Sequence 210764.
9	1.1	663	12	US-10-027-632-180487	Sequence 180487.
10	1.1	663	12	US-10-027-632-180487	Sequence 180487.
11	1.1	663	13	US-10-027-632-180487	Sequence 180487.
12	1.1	663	13	US-10-027-632-180487	Sequence 180487.
13	1.1	1738	14	US-10-204-887-67	Sequence 67, Appli
14	1.1	2000	10	US-09-938-842A-4813	Sequence 4813, Ap
15	1.1	5686	9	US-09-775-938A-31	Sequence 31, Appli
16	1.1	7351	12	US-10-311-455-1	Sequence 1, Appli

17	20	1.1	16950	14	US-10-114-170-166	Sequence 166, App
18	20	1.1	42611	12	US-10-161-127-2	Sequence 2, Appli
19	20	1.1	326014	9	US-09-731-231A-3	Sequence 3, Appli
20	20	1.1	3673778	12	US-10-312-841-1	Sequence 1, Appli
21	19	1.1	394	8	US-08-781-986A-143	Sequence 1943, Ap
22	19	1.1	585	12	US-10-027-632-90502	Sequence 90502, A
23	19	1.1	585	13	US-10-027-632-90502	Sequence 90502, A
24	19	1.1	598	12	US-10-027-632-210749	Sequence 210749.
25	19	1.1	598	13	US-10-027-632-210749	Sequence 210749.
26	19	1.1	644	12	US-10-027-632-190385	Sequence 190385.
27	19	1.1	644	13	US-10-027-632-190385	Sequence 190385.
28	19	1.1	728	12	US-10-027-632-149996	Sequence 149996.
29	19	1.1	728	13	US-10-027-632-149996	Sequence 149996.
30	19	1.1	780	12	US-10-027-632-149997	Sequence 149997.
31	19	1.1	780	13	US-10-027-632-149997	Sequence 149997.
32	19	1.1	813	9	US-09-770-445-763	Sequence 763, App
33	19	1.1	826	14	US-10-091-572-898	Sequence 898, App
34	19	1.1	826	14	US-10-091-572-898	Sequence 900, App
35	19	1.1	1500	10	US-09-822-830A-423	Sequence 423, App
36	19	1.1	1507	12	US-10-027-632-264404	Sequence 264404.
37	19	1.1	1507	13	US-10-027-632-264404	Sequence 264404.
38	19	1.1	5269	12	US-10-311-455-2030	Sequence 2030, Ap
39	19	1.1	5347	12	US-10-311-455-1641	Sequence 1641, App
40	19	1.1	5930	12	US-10-311-455-490	Sequence 490, App
41	19	1.1	6070	12	US-10-204-708-9	Sequence 9, Appli
42	19	1.1	6070	12	US-10-311-455-213	Sequence 213, App
43	19	1.1	6125	12	US-10-311-455-1585	Sequence 1585, App
44	19	1.1	6125	12	US-10-240-453-151	Sequence 151, App
45	19	1.1	6341	12	US-10-311-455-1617	Sequence 1617, Ap

#### ALIGNMENTS

RESULT 1  
US-10-027-632-250060  
; Sequence 250060, Application US/10027632  
; Publication No. US20030204075A9  
; GENERAL INFORMATION:  
; APPLICANT Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027, 632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FASTSeq for Windows Version 4.0  
; SEQ ID NO 250060  
; LENGTH: 1067  
; TYPE: DNA  
; ORGANISM: Human

Query Match 1.2%; Score 22; DB 12; Length 1067;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1037 TGAATTAATGTTTCATGTTA 1058  
|||||

Db 305 TGAATTATGTTTCTATGTTA 326

## RESULT 2

US-10-027-632-250060  
; Sequence 250060, Application US/10027632  
; GENERAL INFORMATION:  
; APPLICANT: Wang, David G.  
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
; FILE REFERENCE: 108827.129  
; CURRENT APPLICATION NUMBER: US/10/027,632  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/218,006  
; PRIOR FILING DATE: 2000-07-12  
; PRIOR APPLICATION NUMBER: US 60/198,676  
; PRIOR FILING DATE: 2000-04-20  
; PRIOR APPLICATION NUMBER: US 60/193,483  
; PRIOR FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: US 60/185,218  
; PRIOR FILING DATE: 2000-02-24  
; PRIOR APPLICATION NUMBER: US 60/167,363  
; PRIOR FILING DATE: 1999-11-23  
; PRIOR APPLICATION NUMBER: US 60/156,358  
; PRIOR FILING DATE: 1999-09-28  
; PRIOR APPLICATION NUMBER: US 60/146,002  
; PRIOR FILING DATE: 1999-08-09  
; NUMBER OF SEQ ID NOS: 325720  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 250060  
; LENGTH: 1067  
; TYPE: DNA  
; ORGANISM: Human  
US-10-027-632-250060

Query Match 1.2%; Score 22; DB 13; Length 1067;  
Best Local Similarity 100.0%; Pred. No. 6.9;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1037 TGAATTATGTTTCTATGTTA 1058  
Db 305 TGAATTATGTTTCTATGTTA 326

## RESULT 3

US-09-960-858-4/c  
; Sequence 4, Application US/09960858  
; Publication No. US2003013877A1  
; GENERAL INFORMATION:  
; APPLICANT: Evans, Glen  
; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF  
; FILE REFERENCE: P-EA 4974  
; CURRENT APPLICATION NUMBER: US/09/960,858  
; CURRENT FILING DATE: 2001-09-20  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 7400  
; TYPE: DNA  
; ORGANISM: M. genitalium  
US-09-960-858-4

Query Match 1.2%; Score 21; DB 12; Length 7400;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1472 AACATTGTTTAACTTGT 1492  
Db 6848 AACATTGTTTAACTTGT 6828

## RESULT 4

US-09-960-870-4/c  
; Sequence 4, Application US/09960870  
; Publication No. US20030134281A1  
; GENERAL INFORMATION:  
; APPLICANT: Evans, Glen  
; TITLE OF INVENTION: NANOMACHINE COMPOSITIONS AND METHODS OF  
; FILE REFERENCE: P-EA 4738  
; CURRENT APPLICATION NUMBER: US/09/960,870  
; CURRENT FILING DATE: 2001-09-20  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 7400  
; TYPE: DNA  
; ORGANISM: M. genitalium  
US-09-960-870-4

Query Match 1.2%; Score 21; DB 12; Length 7400;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1472 AACATTGTTTAACTTGT 1492  
Db 6848 AACATTGTTTAACTTGT 6828

## RESULT 5

US-10-311-455-52  
; Sequence 52, Application US/10311455  
; Publication No. US20030143606A1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIEPENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Detecting  
; FILE REFERENCE: 5013.1014  
; CURRENT APPLICATION NUMBER: US/10/311,455  
; CURRENT FILING DATE: 2002-12-16  
; PRIOR APPLICATION NUMBER: PCT/EP01/07537  
; PRIOR FILING DATE: 2001-07-02  
; PRIOR APPLICATION NUMBER: DE 10032529.7  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: DE 10043826.1  
; PRIOR FILING DATE: 2000-09-01  
; NUMBER OF SEQ ID NOS: 2424  
; SEQ ID NO 52  
; LENGTH: 8011  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
US-10-311-455-52

Query Match 1.2%; Score 21; DB 12; Length 8011;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1359 ATATTATTATATTGTTGT 1379  
Db 5456 ATATTATTATATTGTTGT 5476

## RESULT 6

US-10-205-220-1  
; Sequence 1, Application US/10205220  
; Publication No. US20030170663A1  
; GENERAL INFORMATION:  
; APPLICANT: Fraser et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment  
; FILE REFERENCE: PB1991DI

```

; CURRENT APPLICATION NUMBER: US/10/205,220
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: US 08/545,528
; PRIOR FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/486,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
US-10-205-220-1
```

```

Query Match      1.2%; Score 21; DB 12; Length 580073;
Best Local Similarity 100.0%; Pred. No. 35;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1472 AACATTGGTTTAACTTGT 1492
      |||||||
Db      494176 AACATTGGTTTAACTTGT 494196
```

```

RESULT 7
US-10-027-632-210764
; Sequence 210764, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 210764
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-210764
```

```

Query Match      1.1%; Score 20; DB 12; Length 605;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1036 TTGAATTATGTTTCTATG 1055
      |||||||
Db      240 TTGAATTATGTTTCTATG 259
```

```

RESULT 8
US-10-027-632-210764
; Sequence 210764, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
```

```

; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 210764
; LENGTH: 605
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-210764
```

```

Query Match      1.1%; Score 20; DB 13; Length 605;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1036 TTGAATTATGTTTCTATG 1055
      |||||||
Db      240 TTGAATTATGTTTCTATG 259
```

```

RESULT 9
US-10-027-632-180487
; Sequence 180487, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 180487
; LENGTH: 663
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-180487
```

```

Query Match      1.1%; Score 20; DB 12; Length 663;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1204 CTCCTATTTTAAATTTTACA 1223
```

Db 550 CTTTATTTTAAATTTCACA 569

RESULT 10

US-10-027-632-180488  
Sequence 180488, Application US/10027632  
Publication No. US20030204075A9

GENERAL INFORMATION:

APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/199,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 180488

LENGTH: 663

TYPE: DNA

ORGANISM: Human

US-10-027-632-180488

Query Match

Best Local Similarity 1.1%; Score 20; DB 12; Length 663;  
Best Local Similarity 100.0%; Pred. No. 61;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 CTTTATTTTAAATTTCACA 1223

Db 550 CTTTATTTTAAATTTCACA 569

RESULT 11

US-10-027-632-180487

Sequence 180487, Application US/10027632

GENERAL INFORMATION:

APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/198,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

LENGTH: 663  
TYPE: DNA  
ORGANISM: Human  
US-10-027-632-180487

Query Match

Best Local Similarity 1.1%; Score 20; DB 13; Length 663;  
Best Local Similarity 100.0%; Pred. No. 61;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 CTTTATTTTAAATTTCACA 1223

Db 550 CTTTATTTTAAATTTCACA 569

RESULT 12

US-10-027-632-180488

Sequence 180488, Application US/10027632

GENERAL INFORMATION:

APPLICANT: Wang, David G.  
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide  
Polymorphisms in the Human Genome  
FILE REFERENCE: 108827.129

CURRENT APPLICATION NUMBER: US/10/027,632

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/218,006

PRIOR FILING DATE: 2000-07-12

PRIOR APPLICATION NUMBER: US 60/199,676

PRIOR FILING DATE: 2000-04-20

PRIOR APPLICATION NUMBER: US 60/193,483

PRIOR FILING DATE: 2000-03-29

PRIOR APPLICATION NUMBER: US 60/185,218

PRIOR FILING DATE: 2000-02-24

PRIOR APPLICATION NUMBER: US 60/167,363

PRIOR FILING DATE: 1999-11-23

PRIOR APPLICATION NUMBER: US 60/156,358

PRIOR FILING DATE: 1999-09-28

PRIOR APPLICATION NUMBER: US 60/146,002

PRIOR FILING DATE: 1999-08-09

NUMBER OF SEQ ID NOS: 325720

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 180488

LENGTH: 663

TYPE: DNA

ORGANISM: Human

US-10-027-632-180488

Query Match

Best Local Similarity 1.1%; Score 20; DB 13; Length 663;  
Best Local Similarity 100.0%; Pred. No. 61;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1204 CTTTATTTTAAATTTCACA 1223

Db 550 CTTTATTTTAAATTTCACA 569

RESULT 13

US-10-204-887-67/c

Sequence 67, Application US/10204887

Publication No. US20030124569A1

GENERAL INFORMATION:

APPLICANT: INCTE GENOMICS, INC.  
APPLICANT: PANZER, Scott R.  
APPLICANT: SPIRO, Peter A.  
APPLICANT: BANVILLE, Steven C.  
APPLICANT: SHAH, Purvi  
APPLICANT: CHALUP, Michael S.  
APPLICANT: CHANG, Simon C.  
APPLICANT: CHEN, Alice  
APPLICANT: D'SA, Steven A.  
APPLICANT: AMSHEY, Stefan  
APPLICANT: DAHL, Christopher R.  
APPLICANT: DAM, Tam C.  
APPLICANT: DANIELS, Susan E.

```

; APPLICANT: DUFUR, Gerard E.
; APPLICANT: FLORES, Vincent
; APPLICANT: FONG, Willy T.
; APPLICANT: GREENAWALT, Lila B.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: JONES, Anissa L.
; APPLICANT: LIU, Tommy F.
; APPLICANT: ROSEBERRY, Ann M.
; APPLICANT: ROSEN, Bruce H.
; APPLICANT: RUSSO, Frank D.
; APPLICANT: STOCKREHER, Theresa K.
; APPLICANT: DAFEO, Abel
; APPLICANT: WRIGHT, Rachel J.
; APPLICANT: YAP, Pierre E.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: BRADLEY, Diana L.
; APPLICANT: BRATCHER, Shawn R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: COHEN, Howard J.
; APPLICANT: HODGSON, David M.
; APPLICANT: LINCOLN, Stephen E.
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PT-1134 FCT
; CURRENT APPLICATION NUMBER: US/10/204,887
; PRIOR FILING DATE: 2002-08-21
; PRIOR APPLICATION NUMBER: 60/185,215; 60/185,216; 60/205,232; 60/205,323; 60/205,287;
; 60/205,324; 60/205,286
; PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-05-16; 2000-05-17; 2000-05-17;
; 2000-05-17; 2000-05-17
; NUMBER OF SEQ ID NOS: 159
; SOFTWARE: PERL Program
; SEQ ID NO 67
; LENGTH: 1738
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: incyte ID No. US20030124569A1 LI:336872.1:2000MAY01
US-10-204-887-67

Query Match 1.1%; Score 20; DB 14; Length 1738;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1343 TCTCTCAATATATCTGATAT 1362
DB 620 TCTCTCAATATATCTGATAT 601

RESULT 14
US-09-938-842A-4813/C
; Sequence 4813, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRI1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4813
; LENGTH: 2000
; TYPE: DNA

```

```

; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4813

Query Match 1.1%; Score 20; DB 10; Length 2000;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 AAAGATTATTAACCTGAAT 238
DB 558 AAAGATTATTAACCTGAAT 539

RESULT 15
US-09-775-938A-31
; Sequence 31, Application US/09775938A
; Patent No. US2002008165A1
; GENERAL INFORMATION:
; APPLICANT: Haygood, M.
; APPLICANT: Davidson, S.K.
; APPLICANT: Allen, S.W.
; APPLICANT: Hildebrand, M.
; TITLE OF INVENTION: Bryostatins, Bryopyrans and Polyketides: Compositions and Method
; FILE REFERENCE: 1133.010U1
; CURRENT APPLICATION NUMBER: US/09/775,938A
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: PCT/US00/21326
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/147,283
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 31
; LENGTH: 5686
; TYPE: DNA
; ORGANISM: Endobugula sertula
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(5686)
; OTHER INFORMATION: N refers to any nucleotide.
US-09-775-938A-31

Query Match 1.1%; Score 20; DB 9; Length 5686;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1052 TATGTTACTTTCTTCAGC 1071
DB 4061 TATGTTACTTTCTTCAGC 4060

```

Search completed: November 25, 2003, 20:19:30  
 Job time : 444 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 16:36:37 ; Search time 95 Seconds  
(without alignments)

8284.064 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atccaatctctgattcttga.....tagagcgatcaagctgacc 1783

Scoring table: OLIGO NUC

Gapop 60.0 / Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/PCUS.COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	21	1.2	580073	4 US-08-545-528D-1	Sequence 1, Appl
2	20	1.1	1500	4 US-09-107-532A-1870	Sequence 1870, Ap
3	20	1.1	16950	4 US-09-453-702B-166	Sequence 166, App
4	19	1.1	1882	1 US-08-257-073-12	Sequence 12, Appl
5	19	1.1	1884	1 US-08-257-073-8	Sequence 8, Appl
6	18	1.0	233	2 US-08-687-080-70	Sequence 70, Appl
7	18	1.0	741	4 US-09-328-382-3855	Sequence 3895, Ap
8	18	1.0	1042	3 US-08-765-381-3	Sequence 3, Appl
9	18	1.0	1079	3 US-08-765-381-1	Sequence 1, Appl
10	18	1.0	1303	2 US-08-793-410-11	Sequence 11, Appl
11	18	1.0	1947	4 US-09-134-001C-2324	Sequence 2324, Ap
12	18	1.0	2186	4 US-08-961-527-179	Sequence 179, App
13	18	1.0	3001	4 US-09-539-333D-184	Sequence 184, App
14	18	1.0	83450	4 US-09-811-469-3	Sequence 3, Appl
15	18	1.0	580073	4 US-08-545-528D-1	Sequence 1, Appl
16	17	1.0	40	1 US-08-199-507B-26	Sequence 26, Appl
17	17	1.0	40	1 US-08-441-828-26	Sequence 26, Appl
18	17	1.0	318	3 US-09-018-584A-10	Sequence 10, Appl
19	17	1.0	363	4 US-09-601-198-169	Sequence 169, App
20	17	1.0	471	4 US-09-134-001C-1423	Sequence 1423, Ap
21	17	1.0	683	4 US-09-149-476-263	Sequence 263, App
22	17	1.0	743	4 US-09-220-132-37	Sequence 37, Appl
23	17	1.0	1001	4 US-09-641-638-133	Sequence 133, App
24	17	1.0	1282	4 US-09-475-316A-69	Sequence 69, Appl
25	17	1.0	1320	4 US-09-134-001C-2066	Sequence 2066, Ap
26	17	1.0	1330	4 US-09-252-991A-7869	Sequence 7869, Ap
27	17	1.0	1353	4 US-09-252-991A-7869	Sequence 7869, Ap

28	17	1.0	1398	4 US-09-134-001C-1019	Sequence 1019, Ap
29	17	1.0	1620	4 US-09-328-352-1307	Sequence 1307, Ap
30	17	1.0	1751	4 US-09-149-476-110	Sequence 110, App
31	17	1.0	1939	1 US-07-715-751B-2	Sequence 2, Appl
32	17	1.0	2178	3 US-09-334-601-6	Sequence 6, Appl
33	17	1.0	2288	3 US-09-334-601-1	Sequence 1, Appl
34	17	1.0	2359	4 US-09-425-488-7	Sequence 8, Appl
35	17	1.0	2388	3 US-09-276-531-89	Sequence 43, Appl
36	17	1.0	2404	1 US-08-484-101B-43	Sequence 43, Appl
37	17	1.0	2404	3 US-08-714-524D-43	Sequence 9, Appl
38	17	1.0	2445	3 US-09-298-367B-9	Sequence 152, App
39	17	1.0	2703	4 US-09-620-312D-152	Sequence 7, Appl
40	17	1.0	2883	2 US-08-533-306A-7	Sequence 7, Appl
41	17	1.0	2883	2 US-08-742-923A-7	Sequence 10, Appl
42	17	1.0	2992	2 US-08-841-349-10	Sequence 1597, Ap
43	17	1.0	3267	4 US-09-328-352-1597	Sequence 5, Appl
44	17	1.0	3494	3 US-09-334-601-5	Sequence 1587, Ap
45	17	1.0	3729	4 US-09-107-532A-1587	

## ALIGNMENTS

RESULT 1  
US-08-545-528D-1  
; Sequence 1, Application US/08545528D  
; Patent No. 6537773  
; GENERAL INFORMATION:  
; APPLICANT: Fraser et al.  
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment  
; Patent No. 6537773  
; FILE REFERENCE: PB193PI  
; CURRENT APPLICATION NUMBER: US/08/545,528D  
; CURRENT FILING DATE: 1995-10-19  
; PRIOR APPLICATION NUMBER: US 08/468,018  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/473,545  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 1  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 1  
; LENGTH: 580073  
; TYPE: DNA  
; ORGANISM: Mycoplasma genitalium  
US-08-545-528D-1

Query Match 1.2% Score 21; DB 4; Length 580073;  
Best Local Similarity 100.0%; Pred. No. 1.8;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1472 AACATTGGTTTAACTTGT 1492  
DB 494176 AACATTGGTTTAACTTGT 494196

RESULT 2  
US-09-107-532A-1870  
; Sequence 1870, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stramm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FABRICUM FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02454

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Ariniello, Pamela Denise  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-8277  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 1870:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1500 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: circular  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1..1500  
SEQUENCE DESCRIPTION: SEQ ID NO: 1870:  
US-09-107-532A-1870  
Query Match 1.1% Score 20; DB 4; Length 1500;  
Best Local Similarity 100.0%; Pred. No. 7.5;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 619 GGAGATCTTATGAGAAAT 638  
DB 288 GGAGATCTTATGAGAAAT 307  
RESULT 3  
US-09-453-702B-166  
Sequence 166, Application US/09453702B  
GENERAL INFORMATION:  
APPLICANT: Blatner, Frederick R.  
Perna, Nicole T.  
Blunkett, Guy  
Welch, Rod  
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/453,702B  
FILING DATE: 03-Dec-1999  
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960236.95017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-5166  
INFORMATION FOR SEQ ID NO: 166:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 16950  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 166:  
US-09-453-702B-166  
Query Match 1.1% Score 20; DB 4; Length 16950;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1204 CTTTATTAAATTGACA 1223  
DB 2917 CTTTATTAAATTGACA 2936  
RESULT 4  
US-08-257-073-12/C  
Sequence 12, Application US/08257073  
Patent No. 5765597  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: de Taisne, Charles  
APPLICANT: Tine, John A.  
TITLE OF INVENTION: MALARIA RECOMBINANT POXYVIRUS VACCINE  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue, 25th floor  
CITY: New York  
STATE: New York  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/257,073  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/075,783  
FILING DATE: 11-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/852,305  
FILING DATE: 18-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,183  
FILING DATE: 20-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2570  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066 CURTMS  
INFORMATION FOR SEQ ID NO: 12:



SEQUENCE CHARACTERISTICS:  
LENGTH: 1882 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-257-073-12

Query Match 1.1%; Score 19; DB 1; Length 1882;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1705 TTTTGCATATATTCGTC 1723  
DB 348 TTTTGCATATATTCGTC 330

RESULT 5  
US-08-257-073-8/c  
Sequence 8, Application US/08257073  
Patent No. 5766597  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
APPLICANT: de Taisne, Charles  
TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford, P.C.  
STREET: 530 Fifth Avenue, 25th Floor  
CITY: New York  
STATE: New York  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/257,073  
FILING DATE: 09-JUN-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/075,783  
FILING DATE: 11-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/852,305  
FILING DATE: 18-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/672,183  
FILING DATE: 20-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2570  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
TELEX: 425066 CURTWS  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1884 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-257-073-8

Query Match 1.1%; Score 19; DB 1; Length 1884;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1705 TTTTGCATATATTCGTC 1723  
DB 348 TTTTGCATATATTCGTC 330

DB 348 TTTTGCATATATTCGTC 330

RESULT 6  
US-08-687-080-70  
Sequence 70, Application US/08687080  
Patent No. 5965427

GENERAL INFORMATION:  
APPLICANT: Gregory Dolganov  
TITLE OF INVENTION: Human RAD50 Gene and Methods of Use Thereof  
NUMBER OF SEQUENCES: 175  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Dehlinger & Associates  
STREET: 350 Cambridge Avenue, Suite 250  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94306  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,080  
FILING DATE: 17-JUL-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/592,126  
FILING DATE: 26-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Sholez, Charles K.  
REGISTRATION NUMBER: 38,615  
REFERENCE/DOCKET NUMBER: 4600-0111.30  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 324-0880  
TELEFAX: (415) 324-0960  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 233 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
INDIVIDUAL ISOLATE: INTRON 6 OF RAD50 GENOMIC SEQUENCE  
US-08-687-080-70

Query Match 1.0%; Score 18; DB 2; Length 233;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1379 TCATATTTCTTATGTTT 1396  
DB 182 TCATATTTCTTATGTTT 199

RESULT 7  
US-09-328-352-3895  
Sequence 3895, Application US/09328352  
Patent No. 6562958  
GENERAL INFORMATION:  
APPLICANT: Gary L. Brelton et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER  
TITLE OF INVENTION: BAIWANNI FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC99-032A  
CURRENT APPLICATION NUMBER: US/09/328,352  
CURRENT FILING DATE: 1999-06-04  
NUMBER OF SEQ ID NOS: 8252  
SEQ ID NO 3895  
LENGTH: 741

TYPE: DNA  
ORGANISM: Acinetobacter baumannii  
US-09-328-352-3895

Query Match 1.0%; Score 18; DB 4; Length 741;  
Best Local Similarity 100.0%; Pred. No. 65;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1217 TTTTACACATATATATAA 1234  
DB 62 TTTTACACATATATATAA 79

RESULT 8  
US-08-765-381-3  
Sequence 3, Application US/08765381  
Patent No. 6083724  
GENERAL INFORMATION:  
APPLICANT: Commonwealth Scientific and Industrial Research Organisation  
TITLE OF INVENTION: No. 6083724el avian cytokines and genetic  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESSES:  
ADDRESSES: Scully Scott Murphy and Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City, New York  
STATE: New York  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/765,381  
FILING DATE: 19-DEC-1996  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: AU PN1542/95  
FILING DATE: 06-MAR-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU96/00114  
FILING DATE: 05-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESSER, LEOPOLD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-516-742-4343  
TELEFAX: 1-516-742-4366  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1042 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Chicken (gallus sp.)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 97..588  
US-08-765-381-3

Query Match 1.0%; Score 18; DB 3; Length 1042;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1244 TATTTAAATTTTACAAT 1261  
DB 620 TATTTAAATTTTACAAT 637

RESULT 9  
US-08-765-381-1  
Sequence 1, Application US/08765381  
Patent No. 6083724  
GENERAL INFORMATION:  
APPLICANT: Commonwealth Scientific and Industrial Research Organisation  
TITLE OF INVENTION: No. 6083724el avian cytokines and genetic  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESSES:  
ADDRESSES: Scully Scott Murphy and Presser  
STREET: 400 Garden City Plaza  
CITY: Garden City, New York  
STATE: New York  
COUNTRY: UNITED STATES OF AMERICA  
ZIP: 11530  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/765,381  
FILING DATE: 19-DEC-1996  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: AU PN1542/95  
FILING DATE: 06-MAR-1995  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: PCT/AU96/00114  
FILING DATE: 05-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: PRESSER, LEOPOLD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-516-742-4343  
TELEFAX: 1-516-742-4366  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1079 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
ORIGINAL SOURCE:  
ORGANISM: Chicken (gallus sp.)  
CELL TYPE: T-cell  
CELL LINE: CC8.1h  
IMMEDIATE SOURCE:  
LIBRARY: CC8.1h  
CLONE: CHIFN-gamma  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 134..625  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 191..625  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..133  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 626..1079  
US-08-765-381-1

Query Match 1.0%; Score 18; DB 3; Length 1079;  
Best Local Similarity 100.0%; Pred. No. 64;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1244 TATTTAAATTTTACAAT 1261  
DB 657 TATTTAAATTTTACAAT 674

RESULT 10

```

US-08-793-410-11/C
; Sequence 11, Application US/08793410
; Patent No. 5955650
; GENERAL INFORMATION:
; APPLICANT: HITZ, WILLIAM DEAN
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF CANOLA
; TITLE OF INVENTION: AND SOYBEAN PALMITOYL-ACP THIO-
; TITLE OF INVENTION: ESTERASE GENES AND THEIR USE IN
; TITLE OF INVENTION: THE REGULATION OF FATTY ACID
; TITLE OF INVENTION: CONTENT OF THE OILS OF SOYBEAN
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: USA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; OPERATING SYSTEM: MICROSOFT WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,410
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10627
; FILING DATE: AUGUST 25, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: CHRISTENBURY, LYNN M.
; REGISTRATION NUMBER: 30,971
; REFERENCE/DOCKET NUMBER: CR-9567-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-992-5481
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1303 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; US-08-793-410-11.

Query Match          1.0%; Score 18; DB 2; Length 1303;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      403 TGTGTAATTTGACTT 420
      |||||
DB      577 TGTGTAATTTGACTT 560

RESULT 11
US-09-134-001C-2324/C
; Sequence 2324, Application US/09134001C
; Patent No. 6180370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14

```

```

; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2324
; LENGTH: 1947
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-2324

Query Match          1.0%; Score 18; DB 4; Length 1947;
Best Local Similarity 100.0%; Pred. No. 62;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      544 GTGCAATTGACCATTTTG 561
      |||||
DB      1747 GTGCAATTGACCATTTTG 1730

RESULT 12
US-08-961-527-179/C
; Sequence 179, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 inch, 1.4mb storage
; OPERATING SYSTEM: HP Vectra 486/33
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 179:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2186 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-179

Query Match          1.0%; Score 18; DB 4; Length 2186;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      394 ATTCTAAGTGTAGAA 411
      |||||
DB      466 ATTCTAAGTGTAGAA 449

RESULT 13
US-09-539-333D-184/C
; Sequence 184, Application US/09539333D
; Patent No. 6476208
; GENERAL INFORMATION:
; APPLICANT: Cohen, Daniel

```

```

; APPLICANT: Blumenfeld, Marta
; APPLICANT: Chumakov, Ilya
; APPLICANT: Bouguenel, Lydie
; APPLICANT: Bihain, Bernard
; APPLICANT: Essioux, Laurent
; TITLE OF INVENTION: SCHIZOPHRENIA ASSOCIATED GENES, PROTEINS AND BIALLELIC MARKERS
; FILE REFERENCE: GENSET.047AUS
; CURRENT APPLICATION NUMBER: US/09/539,333D
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: US 60/126,903
; PRIOR FILING DATE: 1999-03-30
; PRIOR APPLICATION NUMBER: US 60/131,971
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/132,065
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: US 60/143,928
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: US 60/145,915
; PRIOR FILING DATE: 1999-07-27
; PRIOR APPLICATION NUMBER: US 60/146,453
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: US 60/146,452
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: US 60/162,288
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: US 09/416,384
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patent.pm
; SEQ ID NO 184
; LENGTH: 3001
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1501
; OTHER INFORMATION: 99-25906-131 : polymorphic base G or T
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1481..1500
; OTHER INFORMATION: 99-25906-131.misl,
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1502..1520
; OTHER INFORMATION: 99-25906-131.mis2, complement
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1374..1392
; OTHER INFORMATION: upstream amplification primer
; FEATURE:
; NAME/KEY: primer_bind
; LOCATION: 1888..1908
; OTHER INFORMATION: downstream amplification primer, complement
; FEATURE:
; NAME/KEY: misc_binding
; LOCATION: 1489..1513
; OTHER INFORMATION: 99-25906-131 probe
; US-09-539-333D-184

Query Match          1.0%; Score 18; DB 4; Length 3001;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1133 AGAATTATGTTATTTCA 1150
Db      778 AGAATTATGTTATTTCA 761

```

```

RESULT 14
US-09-811-469-3/c
; Sequence 3, Application US/09811469
; Patent No. 6551809
; GENERAL INFORMATION:

```

```

; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,
; FILE REFERENCE: C0001171
; CURRENT APPLICATION NUMBER: US/09/811,469
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 83450
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(83450)
; OTHER INFORMATION: n = A,T,C or G
; US-09-811-469-3

Query Match          1.0%; Score 18; DB 4; Length 83450;
Best Local Similarity 100.0%; Pred. No. 49;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1422 AAACGATGATTCATAT 1439
Db      15264 AAACGATGATTCATAT 15247

```

```

RESULT 15
US-08-545-528D-1/c
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide sequence of the Mycoplasma genitalium genome, fragment
; FILE REFERENCE: P019391
; CURRENT APPLICATION NUMBER: US/08/545,528D
; PRIOR FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 580073
; TYPE: DNA
; ORGANISM: Mycoplasma genitalium
; US-08-545-528D-1

Query Match          1.0%; Score 18; DB 4; Length 580073;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1744 TAAACATGACTAATT 1761
Db      574359 TAAACATGACTAATT 574342

```

```

Search completed: November 25, 2003, 19:02:29
Job time : 98 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 15:18:03 ; Search time 367 Seconds  
(without alignments)

13114.708 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783  
Sequence: 1 atccaacttcgtacattga.....tagagcgatcaagctgaacc 1783

Scoring table: OLIGO NUC  
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: N\_Geneseq\_19JUN03: \*  
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: \*  
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: \*  
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: \*  
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: \*  
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: \*  
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: \*  
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: \*  
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: \*  
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: \*  
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: \*  
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: \*  
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: \*  
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: \*  
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: \*  
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: \*  
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: \*  
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: \*  
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: \*  
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: \*  
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: \*  
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: \*  
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: \*  
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: \*  
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1783	100%	1909 21	AAA40223	V. faba sucrose bi
2	21	100%	1909 21	ABL32079	Human immune syste
3	21	100%	8011 24	AA28369	Human chemically t
4	21	100%	112997 24	ABK31320	Signal transductio
5	21	100%	112580073 18	AA758840	Mycoplasma genitali
6	20	100%	994 24	ABX09785	M. incognita RNAi
7	20	100%	11738 22	AA563211	Human purified sec
8	20	100%	1741 24	AB551332	CDNA encoding huma

C	9	20	1.1	2000	24	AB217008
C	10	20	1.1	5686	24	AA229001
C	11	20	1.1	7351	24	ABL32028
C	12	20	1.1	7441	24	ABK40057
C	13	20	1.1	7479	24	AA563344
C	14	20	1.1	7922	24	ABV80086
C	15	20	1.1	32548	23	ABL29482
C	16	20	1.1	42611	25	ABV75791
C	17	20	1.1	129021	21	AAE22296
C	18	20	1.1	326014	24	ABX89296
C	19	19	1.1	394	18	AAV76254
C	20	19	1.1	437	14	AAQ61392
C	21	19	1.1	553	23	ABV57620
C	22	19	1.1	559	23	ABV59186
C	23	19	1.1	579	23	AA567279
C	24	19	1.1	648	21	AA51482
C	25	19	1.1	813	24	ABN98995
C	26	19	1.1	817	21	AA537722
C	27	19	1.1	826	22	AA540746
C	28	19	1.1	826	22	AA540748
C	29	19	1.1	1510	24	AA562636
C	30	19	1.1	1869	13	AAQ29189
C	31	19	1.1	1882	16	AAQ80910
C	32	19	1.1	1882	16	AAQ80913
C	33	19	1.1	2652	10	AAK90703
C	34	19	1.1	3050	25	ABZ10158
C	35	19	1.1	3433	22	AAH41183
C	36	19	1.1	5269	24	ABL34057
C	37	19	1.1	5347	22	ABL33668
C	38	19	1.1	5430	22	AA546291
C	39	19	1.1	5572	22	AA546408
C	40	19	1.1	5930	24	ABL32517
C	41	19	1.1	5947	22	AA546675
C	42	19	1.1	6070	24	ABL92198
C	43	19	1.1	6070	24	ABL49309
C	44	19	1.1	6070	24	ABL32240
C	45	19	1.1	6125	24	ABL33612

## ALIGNMENTS

RESULT 1	AAA40223	standard; DNA; 1909 BP.
AC	AAA40223;	
DT	02-NOV-2000	(first entry)
XX		
DE	V. faba sucrose binding protein DNA fragment.	
XX		
KW	Soybean; sucrose binding protein; SBP; plant seed; transgenic plant;	
KW	seed-specific expression; ds.	
XX		
OS	Vicia faba.	
XX		
XX		
FT	Key	Location/Qualifiers
FT	CDS	1784..1909
FT		/tag= a
FT		/product= "SBP"
FT		/partial
XX		
XX		
XX	WO200026388-A2.	
XX		
XX	11-MAY-2000.	
XX		
XX	27-OCT-1999;	99WO-DE03432.
XX		
XX	04-NOV-1998;	98DE-1052195.
XX		
XX		
XX	(PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.	

Arabidopsis thalia  
Buzula neritina co  
Human immune syste  
Human chemically p  
Chemically pretrea  
Human chemical m  
Diosophila melanog  
Human lipoprotein-  
BAC containing rep  
Human gene for nov  
Staphylococcus aur  
Human brain expres  
Human prostate exp  
Human prostate exp  
DNA encoding novel  
Arabidopsis thalia  
Arabidopsis thalia  
DNA encoding thalia  
DNA encoding human  
CDNA sequence #423  
CDNA encoding plas  
Plasmodium falcipa  
Rhoptry membrane a  
Haematopoietic cel  
Murine G protein-c  
Human immune syste  
Human immune syste  
Tumour suppressor  
Tumour suppressor  
Human immune syste  
Chemically treated  
Human polynucleoti  
Human immune syste  
Human immune syste

PI Heim U, Weber H;  
XX  
XX WPI: 2000-365631/31.  
DR P-PSDB; AAB10028.  
XX  
XX Expression cassette for expressing genes in plant seeds, useful for  
PT producing enzymes or pharmaceuticals, includes the promoter from a  
PT sucrose-binding protein-related gene  
XX  
XX Disclosure; Fig 1; 24pp; German.

XX  
XX This invention describes a novel expression cassette (I) for expressing  
CC genes in plant seeds which comprises (i) the promoter of the  
CC sucrose-binding protein (SBP)-like seed protein (II), (iii) optionally the  
CC sequence for a signal peptide, particularly from (II), (iii) the gene  
CC (iii) to be expressed and (iv) a 3'-termination sequence. (I) and a  
CC plasmid containing the expression cassette (IV) are used for the  
CC expression of homologous or heterologous genes in the seeds of  
CC transformed plants, particularly genes that alter the storage properties  
CC and germination capacity of the seeds. Alternatively, transgenic plants  
CC that express altered or new products in their seeds are selected, grown  
CC to establish stable lines and the resulting products (e.g. enzymes,  
CC pharmaceuticals or proteins that contain essential amino acids) are  
CC extracted. Also insertion of (iii) in the antisense orientation may be  
CC used to reduce or switch off expression of particular genes. (I) provide  
CC seed-specific expression (in cotyledons and endosperm) with stable  
CC expression at high level. They provide an overall increase in expression  
CC rate, improve utilization of the developmental period of the seed and  
CC can overcome the effect of co-suppression. This sequence encodes a  
CC fragment of the Vicia faba (soybean) sucrose binding protein (SBP) which  
CC is described in the method of the invention.

XX  
XX Sequence 1909 BP; 638 A; 317 C; 281 G; 673 T; 0 other;

XX  
XX Query Match 100.0%; Score 1783; DB 21; Length 1909;  
XX Best Local Similarity 100.0%; Pred. No. 0;  
XX Matches 1783; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATCAACCTCTGATCTTGAATCTCTGTCACCAACATGTTGAGAGATTCTAAGACT 60  
DB 1 ATCAACCTCTGATCTTGAATCTCTGTCACCAACATGTTGAGAGATTCTAAGACT 60  
QY 61 TTTCAGAAAGCTTGTAAACATGCTTTGAGACTTTCTTGAATTACTCTTGCACAACTCTGA 120  
DB 61 TTTCAGAAAGCTTGTAAACATGCTTTGAGACTTTCTTGAATTACTCTTGCACAACTCTGA 120  
QY 121 TTGAACTGAGTGAAGAACTGCTCCAGAACTTCAACCAATTCGCTTGGAGAGGCCA 180  
DB 121 TTGAACTGAGTGAAGAACTGCTCCAGAACTTCAACCAATTCGCTTGGAGAGGCCA 180  
QY 121 TTGAACTGAGTGAAGAACTGCTCCAGAACTTCAACCAATTCGCTTGGAGAGGCCA 180  
DB 121 TTGAACTGAGTGAAGAACTGCTCCAGAACTTCAACCAATTCGCTTGGAGAGGCCA 180  
QY 181 AATTTATTTAGTACTTCACTTTCATGAGAGCTGTCTTCAAGATTTTAACTTGAATTC 240  
DB 181 AATTTATTTAGTACTTCACTTTCATGAGAGCTGTCTTCAAGATTTTAACTTGAATTC 240  
QY 241 CATCATTTTAAAGAAAGTTCTGTTCCGCAATGCTTTCATGATTCATTTGAATCTCAACAT 300  
DB 241 CATCATTTTAAAGAAAGTTCTGTTCCGCAATGCTTTCATGATTCATTTGAATCTCAACAT 300  
QY 301 CTGTGTGAGAAAGTTCTTCCAGAACTTGCATCATGTGTAAGAAATCTGGCCAGAAATT 360  
DB 301 CTGTGTGAGAAAGTTCTTCCAGAACTTGCATCATGTGTAAGAAATCTGGCCAGAAATT 360  
QY 361 CTGAACTGTGATATTTCTTAACTGTAAGAAATTTCTAAAGTTTGAATTTGACTT 420  
DB 361 CTGAACTGTGATATTTCTTAACTGTAAGAAATTTCTAAAGTTTGAATTTGACTT 420  
QY 421 TTCCAAAGCAACTGATGATTTGACTTTCTTAATTAACCAACTTCATTTCAACATGT 480  
DB 421 TTCCAAAGCAACTGATGATTTGACTTTCTTAATTAACCAACTTCATTTCAACATGT 480  
QY 481 CTGTGATGAATGTGATTTCTGAATTTGATGTGAGCAAAAGTCAAGATTGACTTTTC 540  
DB 481 CTGTGATGAATGTGATTTCTGAATTTGATGTGAGCAAAAGTCAAGATTGACTTTTC 540

QY 541 AGTGTGAATTTGACATTTTGTCTGTGTCGAATTCGCAACCTTAATTTGATGTGACTG 600  
DB 541 AGTGTGAATTTGACATTTTGTCTGTGTCGAATTTGCAACCTTAATTTGATGTGACTG 600  
QY 601 CTGCAAACTGATGTCTGAGAAAGATCTTATGAGAAATTTCTTGAAGCTGAGAGGAAAA 660  
DB 601 CTGCAAACTGATGTCTGAGAAAGATCTTATGAGAAATTTCTTGAAGCTGAGAGGAAAA 660  
QY 661 TTTTGTGATACAAACAAAGAAATCTGTTTTCCTAATGAGCACTGACATTAACATTA 720  
DB 661 TTTTGTGATACAAACAAAGAAATCTGTTTTCCTAATGAGCACTGACATTAACATTA 720  
QY 721 AACCCACTTATTCGAAAGAGTGAATGAGAGAAATGTCAGTACCTTTCTGCAGTT 780  
DB 721 AACCCACTTATTCGAAAGAGTGAATGAGAGAAATGTCAGTACCTTTCTGCAGTT 780  
QY 781 CATTAAGCAACTTACAGACCTTTTACTTAATTAATCTACAAAGAGGAAATTTTAAACAC 840  
DB 781 CATTAAGCAACTTACAGACCTTTTACTTAATTAATCTACAAAGAGGAAATTTTAAACAC 840  
QY 841 TTAGGAAGTATGGGAGTTTAAAGCAACATTTAAGGGGAGTGTAAATTAATGTG 900  
DB 841 TTAGGAAGTATGGGAGTTTAAAGCAACATTTAAGGGGAGTGTAAATTAATGTG 900  
QY 901 TTGTAAACCACTACCTTTAGTAAATTAAGAAATTTGATATCATCATTAATAT 960  
DB 901 TTGTAAACCACTACCTTTAGTAAATTAAGAAATTTGATATCATCATTAATAT 960  
QY 961 TATTTCTTATTTAAATTTTGAATTAAGTGTATGATTTAGTGAAGAAACCAATATG 1020  
DB 961 TATTTCTTATTTAAATTTTGAATTAAGTGTATGATTTAGTGAAGAAACCAATATG 1020  
QY 1021 TCCTGCTGATATTTTGAATTTTGTCTATGTTACTTTCTTCAAGCCATATPAA 1080  
DB 1021 TCCTGCTGATATTTTGAATTTTGTCTATGTTACTTTCTTCAAGCCATATPAA 1080  
QY 1081 AACTTTGATATGCTAAATTTGATGTGAGAAATTTGATATGATTTCAATTAATTA 1140  
DB 1081 AACTTTGATATGCTAAATTTGATGTGAGAAATTTGATATGATTTCAATTAATTA 1140  
QY 1141 TGGTATTTCAAGTCCAAATTCATCATATGAAATTTAGTCAAAAGCTAACTGAAAAAT 1200  
DB 1141 TGGTATTTCAAGTCCAAATTCATCATATGAAATTTAGTCAAAAGCTAACTGAAAAAT 1200  
QY 1201 ATTCTCTTATTTTAAATTTTACACATATTAATTTCTCTATTTTAAATTTTACAA 1260  
DB 1201 ATTCTCTTATTTTAAATTTTACACATATTAATTTCTCTATTTTAAATTTTACAA 1260  
QY 1261 TAAATATTTTATTCACCTGTGACCTTTGAGAAATTCACCAACATTAATTTAGATAT 1320  
DB 1261 TAAATATTTTATTCACCTGTGACCTTTGAGAAATTCACCAACATTAATTTAGATAT 1320  
QY 1321 TTTATTTCTTAATTTTGAATCTGCAATATATCTGATATTTTATTTATTTTGTGTC 1380  
DB 1321 TTTATTTCTTAATTTTGAATCTGCAATATATCTGATATTTTATTTATTTTGTGTC 1380  
QY 1381 AATTTTCTTATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 1440  
DB 1381 AATTTTCTTATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTTTGAATTT 1440  
QY 1441 TGAATTTTGAAGGACACATTTGACATCTTGAACATTTGATTTTAACTTTGGAATGT 1500  
DB 1441 TGAATTTTGAAGGACACATTTGACATCTTGAACATTTGATTTTAACTTTGGAATGT 1500  
QY 1501 AAAAGTATTAATTAATTTGAGATTTTGAATTTTGAATTTTGAATTTTGAATTT 1560  
DB 1501 AAAAGTATTAATTAATTTGAGATTTTGAATTTTGAATTTTGAATTTTGAATTT 1560  
QY 1561 AAAAGTATTAATTAATTTGAGATTTTGAATTTTGAATTTTGAATTTTGAATTT 1620  
DB 1561 AAAAGTATTAATTAATTTGAGATTTTGAATTTTGAATTTTGAATTTTGAATTT 1620

QY 1621 AATTCATTTCAGATGTAGAACTGCACATACGATTAATTAATTCATTAAGACACGTATG 1680  
 DB 1621 AATTCATTTCAGATGTAGAACTGCACATACGATTAATTAATTCATTAAGACACGTATG 1680  
 QY 1681 TTAACACACGTCCCTTCGATGTTTTTGGCCATAATTCGCTCTCTTTCTTTCTTCAC 1740  
 DB 1681 TTAACACACGTCCCTTCGATGTTTTTGGCCATAATTCGCTCTCTTTCTTTCTTCAC 1740  
 QY 1741 GTATTAACACATGAATTAATTAATGAAGCGATCAAGCTGAACC 1783  
 DB 1741 GTATTAACACATGAATTAATTAATGAAGCGATCAAGCTGAACC 1783

RESULT 2  
 ABL32079  
 ID ABL32079 standard; DNA; 8011 BP.  
 XX  
 AC ABL32079;

26-MAR-2002 (first entry)

Human immune system associated gene SEQ ID NO: 52.

Human; immune system disease; cytosine methylation; antiasthmatic;  
 antiarteriosclerotic; antianaemic; cytosolic; noctropic;  
 neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 antirheumatic; antiarthritic; antidiabetic; ophtalmological;  
 antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 acute myeloid leukemia; Alzheimer's disease; AIDS; epilepsy;  
 neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 gene; ds.

Homo sapiens.

WO200200928-A2.

03-JAN-2002.

02-JUL-2001; 2001WO-EP07537.

30-JUN-2000; 2000DE-1032529.

01-SEP-2000; 2000DE-1043826.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI; 2002-130909/17.

Nucleic acid comprising fragment of chemically modified gene, useful

for diagnosis and treatment of diseases associated with abnormal

cytosine methylation

Claim 1; SEQ ID NO 52; 32pg + Sequence Listing; German.

The present invention provides a number of human immune system associated  
 genes which are modified by the methylation of cytosines. The sequences  
 can be used in the diagnosis and treatment of immune system disorders,  
 including eye diseases such as retinopathy, neovascular glaucoma and  
 macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 diseases. The present sequence is a gene of the invention.

Sequence 8011 BP; 2439 A; 145 C; 1810 G; 3617 T; 0 other;

Query Match 1.2%; Score 21; DB 24; Length 8011;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1359 ATATTATTATTATTATTGTTGTT 1379  
 DB 5456 ATATTATTATTATTATTGTTGTT 5476

RESULT 3  
 AAD28369  
 ID AAD28369 standard; DNA; 8011 BP.  
 XX  
 AC AAD28369;

22-APR-2002 (first entry)  
 Human chemically treated genomic DNA #10.

Human; cytosolic; antidepressant; neuroleptic; noctropic; antiaddictive;  
 adrenergic alpha-1C-receptor; cytosine methylation; therapy; alcoholism;  
 behavioural disorder; neurological; psychiatric; cancer; schizophrenia;  
 Tourette's syndrome; smoking; human immunodeficiency virus dementia;  
 drug abuse; migraine; ds.

Homo sapiens.

WO200202809-A2.

10-JAN-2002.

02-JUL-2001; 2001WO-EP07540.

30-JUN-2000; 2000DE-1032529.

01-SEP-2000; 2000DE-1043826.

(EPIG-) EPIGENOMICS AG.

Olek A, Piepenbrock C, Berlin K;

WPI; 2002-154759/20.

Novel nucleic acid useful for diagnosis and therapy of behavioral

disorder, neurological disorder and cancer, comprises a sequence of a

segment of chemically pretreated DNA of adrenergic alpha-1C-receptor

gene

Claim 1; Page 60-64; 190pp; English.

The invention relates to nucleic acids comprising a segment of chemically  
 pretreated DNA of adrenergic alpha-1C-receptor gene. The invention also  
 relates to oligonucleotides or peptide nucleic acid (PNA) oligomers  
 useful for detecting cytosine methylations. The pretreated DNA is useful  
 for the diagnosis or therapy of behavioural disorders, neurological  
 disorders and cancer, in particular major depressive disorder, Tourette's  
 syndrome, schizophrenia, psychiatric and neurological disorders, smoking,  
 drug abuse, alcoholism, personality traits, compulsive gambling, human  
 immunodeficiency virus dementia, migraine, behaviours in schizophrenia  
 and schizophrenia. The nucleic acid is useful for detecting the methylation  
 state of all CpG dinucleotides and/or single nucleotide polymorphisms  
 (SNPs). The present sequence is human chemically treated genomic DNA.

Sequence 8011 BP; 2439 A; 145 C; 1810 G; 3617 T; 0 other;

Query Match 1.2%; Score 21; DB 24; Length 8011;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1359 ATATTATTATTATTATTGTTGTT 1379  
 DB 5456 ATATTATTATTATTATTGTTGTT 5476

RESULT 4  
 ABL31320  
 ID ABL31320 standard; DNA; 11907 BP.  
 XX  
 AC ABL31320;  
 XX

DT 23-APR-2002 (first entry)  
 XX  
 KW Signal transduction associated gene modified DNA #82.  
 DE  
 XX  
 KW Human, signal transduction associated gene; cytosine methylation state;  
 KW CpG island; signal transduction associated disease; solid tumour; cancer;  
 KW antitumour; cytostatic; mutant; ds.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FN WO200200926-A2.  
 PD  
 XX 03-JAN-2002.  
 PD  
 XX 29-JUN-2001; 2001WO-EP07472.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-147896/19.  
 XX  
 PT Oligonucleotide for diagnosis and therapy of diseases associated with  
 PT signal transduction e.g. cancer, comprises chemically modified genomic  
 PT sequences of genes associated with signal transduction -  
 XX  
 PS Claim 1; SEQ ID No 163; 24pp; English.  
 XX  
 CC The present invention relates to chemically modified DNA sequences of  
 CC signal transduction associated genes. The DNA sequences are chemically  
 CC modified using a solution of bisulphite, hydrogen sulphite or  
 CC disulphite. Also disclosed are oligonucleotides and/or PNA oligomers  
 CC for detecting the cytosine methylation state (CpG islands) of these  
 CC genes, and a method for the diagnosis and/or therapy of genetic and  
 CC epigenetic parameters of genes associated with signal transduction.  
 CC The genomic DNA can be obtained from cells or cellular components which  
 CC contain DNA, e.g. cell lines, biopsies, blood, sputum, stool, urine,  
 CC cerebrospinal fluid, tissue embedded in paraffin such as tissue from  
 CC eye, intestine, kidney, brain, heart, prostate, lung, breast or liver,  
 CC histologic object slides, and all their possible combinations. The  
 CC sequences of the invention are useful for the diagnosis and therapy of  
 CC diseases associated with signal transduction e.g. solid tumours and  
 CC cancer. ABX31158-ABX31945 represent chemically pretreated genomic DNA  
 CC sequences of different genes associated with signal transduction, or  
 CC their complementary sequences.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 SQ Sequence 11907 BP; 3442 A; 180 C; 2430 G; 5855 T; 0 other;  
 Query Match 1.2%; Score 21; DB 24; Length 11907;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1359 ATATTATTATTTATTTGCT 1379  
 Db 2943 ATATTATTATTTATTTGCT 2963  
 RESULT 5  
 AAT58840  
 ID AAT58840 standard; DNA; 580073 BP.  
 XX  
 AC AAT58840;  
 XX  
 DT 27-MAR-1997 (first entry)  
 XX  
 DE \*Mycoplasma genitalium genome.

XX M. genitalium; DNA; DNA gyrase; origin of replication;  
 KW megabase shotgun sequencing method; open reading frame; ORF; ss.  
 KW  
 XX  
 XX Mycoplasma genitalium.  
 PH  
 FT Key  
 FT CDS  
 FT Location/Qualifiers  
 FT 8552..9184  
 FT /tag= a  
 FT /label= MG006  
 FT /note= "Previously identified as MORF-20076, the  
 FT encoded protein shows 27.59 percentage  
 FT identity to thymidylate kinase (CDC8)  
 FT from Saccharomyces cerevisiae"  
 FT 11252..12040  
 FT /tag= b  
 FT /label= MG009  
 FT /note= "Previously identified as MORF-20078, the  
 FT encoded protein shows 35.43 percentage  
 FT identity to the Bacillus subtilis hypothetical  
 FT protein covered in accession number  
 FT GB:D26185\_102"  
 FT 12069..12725  
 FT /tag= c  
 FT /label= MG010  
 FT /note= "Previously identified as MORF-20079, the  
 FT encoded protein shows 25.73 percentage  
 FT identity to DNA primase (dnaE) from  
 FT Clostridium acetobutylicum"  
 FT complement (13570..14247)  
 FT /tag= d  
 FT /label= MG012  
 FT /note= "Previously identified as MORF-20080, the  
 FT encoded protein shows 31.50 percentage  
 FT identity to the ribosomal protein S6  
 FT modification protein (rimK) from Escherichia  
 FT coli"  
 FT complement (14396..15217)  
 FT /tag= e  
 FT /label= MG013  
 FT /note= "Previously identified as MORF-19823, MORF-20080  
 FT and MORF-20081, the encoded protein shows 33.04  
 FT percentage identity to 5,10-methylene-tetra-  
 FT hydrofolate dehydrogenase (folD) from E. coli"  
 FT 17474..19243  
 FT /tag= f  
 FT /label= MG015  
 FT /note= "Previously identified as MORF-20084, the  
 FT encoded protein shows 32.23 percentage  
 FT identity to transport ATP-binding protein  
 FT (msbA) from E. coli"  
 FT 26478..27344  
 FT /tag= g  
 FT /label= MG023  
 FT /note= "Previously identified as MORF-20092, the  
 FT encoded protein shows 45.96 percentage  
 FT identity to fructose-bisphosphate aldolase  
 FT (tss) from B. subtilis"  
 FT 27345..28448  
 FT /tag= h  
 FT /label= MG024  
 FT /note= "Previously identified as MORF-19826 and  
 FT MORF-20093, the encoded protein shows 46.84  
 FT percentage identity to GTP-binding protein  
 FT from E. coli"  
 FT 36987..38978  
 FT /tag= i  
 FT /label= MG032  
 FT /note= "Previously identified as MORF-20099, the  
 FT encoded protein shows 26.82 percentage  
 FT identity to ATP-dependent nuclease (addA)  
 FT from B. subtilis"  
 FT 39242..39904  
 FT CDS



```

FT      /tag= j
FT      /label= MG033
FT      /note= "Previously identified as MORF-20100, the
FT      encoded protein shows 35.90 percentage
FT      identity to glycerol uptake facilitator
FT      (glpF) from B. subtilis"
FT      CDS
FT      complement (39873..40514)
FT      /tag= x
FT      /label= MG034
FT      /note= "Previously identified as MORF-20101, the
FT      encoded protein shows 48.13 percentage
FT      identity to thymidylate kinase (tdk)
FT      from B. subtilis"
FT      CDS
FT      40543..41787
FT      /tag= l
FT      /label= MG035
FT      /note= "Previously identified as MORF-20102, the
FT      encoded protein shows 30.71 percentage
FT      identity to histidyl-tRNA synthetase (hiss)
FT      from Mycobacterium leprae"
FT      CDS
FT      complement (44751..46277)
FT      /tag= m
FT      /label= MG038
FT      /note= "Previously identified as MORF-20105, the
FT      encoded protein shows 46.83 percentage
FT      identity to glycerol kinase (glpK)
FT      from E. coli"
FT      CDS
FT      complement (46268..47422)
FT      /tag= n
FT      /label= MG039
FT      /note= "Previously identified as MORF-19831 and
FT      MORF-20106, the encoded protein shows 43.20
FT      percentage identity to glycerol-3-phosphate
FT      dehydrogenase (GDH2) from S. cerevisiae"
FT      CDS
FT      49377..49643
FT      /tag= o
FT      /label= MG041
FT      /note= "The encoded protein shows 48.86 percentage
FT      identity to phosphohistidinoprotein-hexose
FT      phosphotransferase (pghs) from Mycoplasma
FT      capricolum"
FT      CDS
FT      50060..51520
FT      /tag= p
FT      /label= MG042
FT      /note= "Previously identified as MORF-19832 and
FT      MORF-20108, the encoded protein shows 41.92
FT      percentage identity to spermidine/
FT      putrescine transport ATP-binding protein
FT      (potA) from E. coli"
FT      CDS
FT      51525..52382
FT      /tag= q
FT      /label= MG043
FT      /note= "Previously identified as MORF-20110, the
FT      encoded protein shows 26.51 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein (potB) from E. coli"
FT      CDS
FT      52366..53220
FT      /tag= r
FT      /label= MG044
FT      /note= "Previously identified as MORF-20111, the
FT      encoded protein shows 29.45 percentage
FT      identity to spermidine/putrescine transport
FT      system permease protein C (potC) from E. coli"
FT      CDS
FT      54658..55605
FT      /tag= s
FT      /label= MG046
FT      /note= "Previously identified as MORF-20112, the
FT      encoded protein shows 36.60 percentage
FT      identity to stiaoglycoprotease (gcp)
FT      from Pasteurella haemolytica"
FT      CDS
FT      complement (56970..58310)
FT      /tag= t
FT      /label= MG048
FT      /note= "Previously identified as MORF-19834,
FT      MORF-20114 and MORF-20115, the encoded protein
FT      shows 43.02 percentage identity to signal
FT      recognition particle protein (fth) from B.
FT      subtilis"
FT      CDS
FT      58117..53079
FT      /tag= u
FT      /label= MG049
FT      /note= "Previously identified as MORF-20114 and
FT      MORF-20115, the encoded protein shows 44.78
FT      percentage identity to purine-nucleoside
FT      phosphorylase (deod) from E. coli"
FT      CDS
FT      59083..59754
FT      /tag= v
FT      /label= MG050
FT      /note= "Previously identified as MORF-20117, the
FT      encoded protein shows 83.03 percentage
FT      identity to deoxyribose-phosphate aldolase
FT      (deac) from Mycoplasma pneumoniae"
FT      CDS
FT      complement (64898..65731)
FT      /tag= w
FT      /label= MG056
FT      /note= "Previously identified as MORF-20122, the
FT      encoded protein shows 30.25 percent
FT      identity to the protein disclosed in
FT      GB:D26185_99 from B. subtilis"
FT      CDS
FT      complement (65713..66249)
FT      /tag= x
FT      /label= MG057
FT      /note= "Previously identified as MORF-20123, the
FT      encoded protein shows 38.90 percentage
FT      identity to the protein disclosed in
FT      GB:D26185_104 from B. subtilis"
FT      CDS
FT      81047..82597
FT      /tag= y
FT      /label= MG067
FT      /note= "Previously identified as MORF-19845, the
FT      encoded protein shows 28.84 percentage
FT      identity to glutamic acid specific protease
FT      (SPase) from Staphylococcus aureus"
FT      CDS
FT      91065..91919
FT      /tag= z
FT      /label= MG070
FT      /note= "Previously identified as MORF-20136, the
FT      encoded protein shows 34.8 percentage
FT      identity to ribosomal protein S2 (rps2)
FT      from Spirulina plantensis"
FT      Query Match 1.2%; Score 21; DB 18; Length 580073;
FT      Best Local Similarity 100.0%; Pred. No. 13;
FT      Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1472 AACATGCTTTAACTTGT 1492
DB 494176 AACATTGTTTAACTTGT 494196

```

## RESULT 6

ABX09785

ID ABX09785 standard; DNA; 994 BP.

AC ABX09785;

XX 22-JAN-2003 (first entry)

DB M. incognita RNAi molecule #42 useful for nematode control.

XX RNAi molecule; double-stranded interfering RNA; nematode control;

XX growth; development; parasitism; reproduction; RNAi vector;

XX anti-nematode; ds.

OS Meloidogyne incognita.  
 XX  
 XX WO200196584-A2.  
 XX  
 XX 20-DEC-2001.  
 PD  
 XX 12-JUN-2001; 2001WO-US18911.  
 XX  
 XX 12-JUN-2001; 2001WO-US18911.  
 XX  
 XX 12-JUN-2000; 2000US-210917P.  
 PR  
 XX (AKKA-) AKKADIX CORP.  
 PA  
 XX Mushegian AR, Taylor CG, Feltelson JS, Eroshkin AM;  
 PI WPI; 2002-139714/18.  
 DR  
 XX RNA mediated interference molecule useful for disrupting cellular  
 PT process in a nematode, for controlling nematodes comprises genetic  
 PT regulatory sequences -  
 XX  
 XX Claim 1; Page 96; 103pp; English.  
 PS  
 XX The present invention relates to RNAi (double-stranded interfering  
 CC RNA or RNA mediated interference) molecules (nematode genes), and  
 CC methods of using these sequences in nematode control. RNAi molecules  
 CC selectively target mRNA transcripts of essential nematode genes. The  
 CC RNAi molecules of the invention are useful for disrupting cellular  
 CC processes in a nematode by contacting the nematode with a composition  
 CC comprising an RNAi molecule. The RNAi molecules are useful for killing  
 CC nematodes and/or inhibiting their growth, development, parasitism or  
 CC reproduction and also for the regulation of levels of specific RNA  
 CC in nematodes. Genetic regulatory sequences such as promoters, enhancers  
 CC and terminators can be used in genetic constructs such as RNAi vectors  
 CC which can be used for nematode control. The RNAi molecules are  
 CC capable of targeting and reducing (and, in some cases, preventing)  
 CC the translation of a specific gene product, and can be used to reduce  
 CC or prevent mRNA translation in any tissue of the nematode because of  
 CC its ability to cross tissue and cellular boundaries. The RNAi molecule  
 CC can be contacted with a nematode by soaking, injection, or consumption  
 CC of a food source containing an RNAi molecule. The RNAi molecules can  
 CC also be used as an epigenetic factor to prevent the proliferation of  
 CC subsequent generations of nematodes, to produce nematode inhibitors  
 CC or RNAi in the plants, and provide new biotechnological strategies for  
 CC managing nematodes under sustainable agricultural conditions.  
 CC ABX09677-ABX09815 represent RNAi molecules useful for the control of  
 CC nematodes.  
 CC  
 XX  
 SQ Sequence 994 BP; 357 A; 155 C; 180 G; 302 T; 0 other;  
 Query Match 1.1%; Score 20; DB 24; Length 994;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 943 TAATCATCATTTAATTA 962  
 DB 21 TAATCATCATTTAATTA 40

XX Homo sapiens.  
 OS  
 XX WO200162918-A2.  
 XX  
 XX 30-AUG-2001.  
 PD  
 XX 01-FEB-2001; 2001WO-US03465.  
 XX  
 XX 24-FEB-2000; 2000US-185215P.  
 XX  
 XX 24-FEB-2000; 2000US-185216P.  
 PR 16-MAY-2000; 2000US-205232P.  
 PR 17-MAY-2000; 2000US-205286P.  
 PR 17-MAY-2000; 2000US-205287P.  
 PR 17-MAY-2000; 2000US-205323P.  
 PR 17-MAY-2000; 2000US-205324P.  
 XX  
 XX (INCYTE) INCYTE GENOMICS INC.  
 PA  
 XX Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;  
 PI Chen A, D'sa SA, Amshay S, Dahl CR, Dam TC, Daniels SE;  
 PI Dufour GE, Flores V, Fong WT, Greenwalt LB, Hillman JL, Jones AL;  
 PI Liu TF, Roseberry AM, Rosen BH, Russo FD, Stockbreher TK, Daffo A;  
 PI Wright RJ, Yap PE, Yu YJ, Bradley DL, Bratcher SR, Chen W;  
 PI Cohen RJ, Hodgson DW, Lincoln SE;  
 DR WPI; 2001-648217/74.  
 XX  
 XX Nucleic acids encoding secretory polypeptides, useful in genetic  
 PT diagnosis and therapy -  
 PT  
 XX Claim 1; Page 181-182; 237pp; English.  
 PS  
 XX Sequences AAS63145-AAS63223 represent DNA encoding purified secretory  
 CC polypeptides of the invention. The polypeptides and polynucleotides can  
 CC be used in the treatment, prevention and diagnosis of diseases associated  
 CC with inappropriate secretory protein expression. These diseases include  
 CC cell proliferative disorders such as atherosclerosis and psoriasis,  
 CC cancers such as leukemia and melanoma, immune system disorders such as  
 CC asthma and diabetes mellitus, neurological disorders such as epilepsy and  
 CC parkinson's disease, mental disorders such as schizophrenia and seasonal  
 CC affective disorder (SAD), motor neuron disorders such as amyotrophic  
 CC lateral sclerosis, demyelinating disorders such as multiple sclerosis,  
 CC central nervous system disorders such as mental retardation and  
 CC neurofibromatosis and neuromuscular disorders such as cerebral palsy and  
 CC muscular dystrophy. Target polynucleotides in a sample can be detected by  
 CC hybridizing the sample with a probe sequence complementary to the target  
 CC polynucleotide, under conditions in which a hybridisation complex is  
 CC formed, and detecting the presence or absence of the complex. The  
 CC polypeptides may also be used as antigens in the production of antibodies  
 CC against secretory proteins and in assays to identify modulators of  
 CC protein expression and activity. The antibodies may also be used as  
 CC diagnostic agents for detecting the presence of the sequences of the  
 CC invention in samples e.g. by enzyme linked immunosorbent assay (ELISA).  
 CC  
 XX  
 SQ Sequence 1738 BP; 559 A; 357 C; 332 G; 490 T; 0 other;  
 Query Match 1.1%; Score 20; DB 22; Length 1738;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1343 TCTCTCAATATATCGAAT 1362  
 DB 620 TCTCTCAATATATCGAAT 601

RESULT 7  
 AAS63211/c  
 ID AAS63211 standard; cDNA; 1738 BP.  
 XX  
 XX AAS63211;  
 XX  
 XX 29-JAN-2002 (first entry)  
 XX  
 XX Human purified secretory polynucleotide #67.  
 DE  
 XX Human; purified secretory polypeptide; cell proliferative disorder; ss;  
 KM cancer; immune system disorder; neurological disorder; mental disorder;  
 KM motor neuron disorder; demyelinating disorder; neuromuscular disorder;  
 KM central nervous system disorder; enzyme linked immunosorbent assay;  
 KM ELISA; gene therapy.

RESULT 8  
 AAS51332/c  
 ID AAS51332 standard; cDNA; 1741 BP.  
 XX  
 XX AAS51332;  
 AC  
 XX 21-OCT-2002 (first entry)  
 DT

XX cDNA encoding human secretory protein #30.

DE Human; secretory polypeptide; SPM; actinic keratosis; arteriosclerosis;  
 KW buritis; cirrhosis; hepatitis; polycythaemia vera; anaemia; psoriasis;  
 KW primary thrombocytopenia; cancer; adenocarcinoma; leukaemia; myeloma;  
 KW sarcoma; immune system disorder; acquired immunodeficiency syndrome;  
 KW AIDS; allergy; asthma; Crohn's disease; diabetes mellitus; gout;  
 KW glomerulonephritis; Goodpasture's syndrome; thyroiditis; pancreatitis;  
 KW hepatitis; multiple sclerosis; osteoporosis; Reiter's syndrome;  
 KW rheumatoid arthritis; neurological disorder; epilepsy; stroke; dementia;  
 KW Alzheimer's disease; Pick's disease; Huntington's disease; mood; anxiety;  
 KW Parkinson's disease; central nervous system disorder; mental disorder;  
 KW schizophrenic disorder; amnesia; Tourette's disorder; transgenic animal;  
 KW gene therapy; gene; ss.

OS Homo sapiens.

XX WO200257304-A2.

PN 25-JUL-2002.

XX 15-JAN-2002; 2002WO-US01340.

XX 16-JAN-2001; 2001US-261864P.  
 PR 16-JAN-2001; 2001US-261865P.  
 PR 16-JAN-2001; 2001US-261979P.  
 PR 16-JAN-2001; 2001US-261981P.  
 PR 17-JAN-2001; 2001US-262164P.  
 PR 17-JAN-2001; 2001US-262208P.  
 PR 17-JAN-2001; 2001US-263131P.  
 PR 19-JAN-2001; 2001US-263598P.  
 PR 19-JAN-2001; 2001US-263780P.  
 PR 19-JAN-2001; 2001US-263063P.  
 PR 19-JAN-2001; 2001US-263066P.  
 PR 19-JAN-2001; 2001US-263069P.  
 PR 19-JAN-2001; 2001US-263070P.  
 PR 19-JAN-2001; 2001US-263074P.  
 PR 19-JAN-2001; 2001US-263076P.  
 PR 19-JAN-2001; 2001US-263077P.  
 PR 19-JAN-2001; 2001US-263329P.

XX (INCY-) INCYTE GENOMICS INC.

XX Panzer SR, Lincoln SE, Altus CM, Dufour GE, Hillman JL, Jones AL;  
 PI Dam TC, Liu TF, Harris B, Flores V, Daffo A, Marwaha R, Chen AJ;  
 PI Chang SC, Gerstein EH, Peralta CH, David MR, Lewis SA;  
 DR WPI; 2002-590716/63.  
 DR N-PSDB; ABG69840, ABG69841.

XX New purified secretory polypeptides and polynucleotides, useful in the  
 PT diagnosis, study, prevention or treatment of diseases associated with  
 PT decreased expression of functional secretory molecules, e.g. AIDS,  
 PT cancer or allergies

XX Claim 1; Page 266; 340P; English.

XX The invention describes an isolated polynucleotide a naturally occurring  
 CC polynucleotide sequence at least 90 % identical to it, a polynucleotide  
 CC complementary to it or an RNA equivalent of it. The purified secretory  
 CC polypeptides (SPM) and polynucleotides are useful in the diagnosis,  
 CC study, prevention or treatment of diseases associated with decreased  
 CC expression of functional SPM, e.g. actinic keratosis, arteriosclerosis,  
 CC buritis, cirrhosis, hepatitis, polycythaemia vera, primary  
 CC thrombocytopenia, anaemia, psoriasis, cancers including adenocarcinoma,  
 CC leukemia, myeloma or sarcoma, immune system disorder such as acquired  
 CC immunodeficiency syndrome (AIDS), allergies, asthma, Crohn's disease,  
 CC diabetes mellitus, glomerulonephritis, Goodpasture's syndrome, gout,  
 CC Hashimoto's thyroiditis, hepatitis, multiple sclerosis, osteoporosis,  
 CC pancreatitis, Reiter's syndrome, autoimmune thyroiditis or rheumatoid  
 CC arthritis, neurological disorders such as epilepsy, stroke, Alzheimer's  
 CC disease, Pick's disease, Huntington's disease, dementia, Parkinson's

CC disease, other developmental disorder of the central nervous system,  
 CC mental disorder including mood, anxiety or schizophrenic disorder,  
 CC amnesia or Tourette's disorder. The polynucleotides may be used in  
 CC hybridisation and amplification technologies, e.g. in assessing gene  
 CC expression patterns, to develop a transcript image for a particular cell  
 CC or tissue, or to create transgenic animals to model human disease. This  
 CC sequence encodes a human secretory protein isolated in the invention.

XX SQ Sequence 1741 BP; 560 A; 357 C; 333 G; 491 T; 0 other;

XX Query Match 1.1%; Score 20; DB 24; Length 1741;  
 XX Best Local Similarity 100.0%; Pred. No. 49;  
 XX Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX Db 1343 TCTCTCAATATATCTGATAT 1362  
 XX 622 TCTCTCAATATATCTGATAT 603

XX RESULT 9  
 XX AB217008/c  
 XX ID AB217008 standard; DNA; 2000 BP.

XX AC AB217008;  
 XX DT 21-JAN-2003 (first entry)

XX DE Arabidopsis thaliana stress regulated gene SEQ ID NO 4813.

XX KW Arabidopsis thaliana; plant; gene; stress; transgenic; ds.

XX OS Arabidopsis thaliana.

XX PN WO200216655-A2.  
 XX PD 28-FEB-2002.

XX PF 24-AUG-2001; 2001WO-US26685.  
 XX PR 24-AUG-2000; 2000US-227866P.  
 XX PR 26-JAN-2001; 2001US-264647P.  
 XX PR 22-JUN-2001; 2001US-300111P.

XX PA (SCR1) SCRIPPS RES. INST.  
 XX PA (SYGN) SYNGENTA PARTICIPATIONS AG.  
 XX PI Harper JF, Kreps J, Wang X, Zhu T;  
 DR WPI; 2002-304127/34.

XX PT Identifying a stress condition to which a plant cell has been exposed  
 PT and producing plants with increased tolerance to these abiotic stresses

XX Claim 14; SEQ ID NO 4813; 577bp + Sequence Listing; English.

XX The invention relates to identifying a stress condition to which a plant  
 CC cell has been exposed, comprising:  
 CC (a) contacting nucleic acid representative of expressed polynucleotides  
 CC in the plant cell with an array or probes representative of the plant  
 CC cell genome; and  
 CC (b) detecting a profile of expressed polynucleotides in the plant cell  
 CC characteristic of a stress response. The method is useful in the  
 CC production of transgenic plants, cells and seeds and in producing plants  
 CC with increased tolerance to abiotic stress. The present sequence is that  
 CC of an Arabidopsis thaliana stress regulated gene (AB212196-AB217574) used  
 CC in methods of the invention.  
 CC Note: The sequence data for this patent is not represented in the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office.

XX SQ Sequence 2000 BP; 723 A; 323 C; 280 G; 674 T; 0 other;

Query Match 1.1%; Score 20; DB 24; Length 2000;  
 Best Local Similarity 100.0%; Pred. No. 49;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

219 AAGATTATTAAGTGAAT 238  
 558 AAGATTATTAAGTGAAT 539

RESULT 10  
 AAD29001  
 ID AAD29001 standard; DNA; 5686 BP.  
 XX  
 AC AAD29001;  
 XX  
 DT 07-MAY-2002 (first entry)  
 XX  
 DE Bugula neritina contig 5 DNA sequence from cosmid 6A.  
 XX  
 KW Polyketide; bryopyran ring; byrostatin; breast cancer; anticancer;  
 KW antifungal; antimicrobial; immunomodulatory; polyketide synthase;  
 KW PKS; enzyme; ds.  
 XX  
 OS Bugula neritina.  
 XX  
 OS WO20011024-A2.  
 XX  
 PD 15-FEB-2001.  
 XX  
 PF 04-AUG-2000; 2000WO-US21326.  
 XX  
 PR 04-AUG-1999; 99US-147283P.  
 XX  
 XX (REGC) UNIV CALIFORNIA.  
 XX  
 PI Haygood M, Davidson SK, Allen SW, Hildebrand M;  
 XX  
 DR WPI; 2002-154285/20.  
 XX  
 PT Composition comprising a polypeptide isolated from marine organism,  
 PT which catalyzes at least one step in synthesis of polyketide/bryopyran  
 PT ring, useful for producing polyketide or bryopyran ring containing  
 PT compositions -  
 XX  
 PS Claim 64; Fig 15B; 233p; English.  
 XX  
 CC The invention relates to compositions comprising nucleic acid molecules  
 CC encoding a polypeptide which catalyzes at least one step in synthesis of  
 CC polyketides including bryopyran ring, such as byrostatin. These novel  
 CC sequences are derived from marine organisms. Compositions containing  
 CC sequences of the invention are useful for producing base structure,  
 CC bryopyran rings that can form the basis of combinatorial chemistry to  
 CC form a wide variety of compounds which can be screened for bioactivities  
 CC including anticancer activity. The cloned genes and linked genes  
 CC involved in byrostatin synthesis can be used to screen environmental  
 CC samples for polyketide synthase (PKS) genes. They are also used for  
 CC combinatorial creation of novel polyketide/byrostatin analogues that may  
 CC exhibit improved anti-cancer properties. Compositions of the invention  
 CC are useful for producing byrostatin and its analogues which are useful  
 CC for treating breast cancer and as anticancer, antifungal antimicrobial  
 CC and immunomodulatory compounds. They are useful for producing novel  
 CC polyketides such as bryopyran rings including byrostatin. The present  
 CC sequence is Bugula neritina contig 5 DNA sequence from cosmid 6A  
 CC used in the exemplification of the invention.  
 CC  
 SO Sequence 5686 BP; 1601 A; 924 C; 1195 G; 1514 T; 452 other;  
 XX  
 SQ  
 Query Match 1.1%; Score 20; DB 24; Length 5686;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1052 TATGTTACTTTCTTCAAGC 1071

DB 4061 TATGTTACTTTCTTCAAGC 4060

RESULT 11  
 ABL32028  
 ID ABL32028 standard; DNA; 7351 BP.  
 XX  
 AC ABL32028;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 1.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antihaemic; cytosolic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;  
 KW gene; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 OS WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07537.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 XX  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful  
 PT for diagnosis and treatment of diseases associated with abnormal  
 PT cytosine methylation -  
 XX  
 PS Claim 1; SEQ ID NO 1; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention.  
 CC  
 SO Sequence 7351 BP; 2230 A; 42 C; 1230 G; 3849 T; 0 other;  
 XX  
 SQ  
 Query Match 1.1%; Score 20; DB 24; Length 7351;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

969 TTATTTAAATATGATGATAA 988  
 6500 TTATTTAAATATGATGATAA 6519

RESULT 12  
 ABL40057/C  
 ID ABL40057 standard; DNA; 7441 BP.  
 XX  
 AC ABL40057;  
 XX  
 DT 21-MAY-2002 (first entry)  
 XX

DE Human chemically pretreated gene sequence #70 strand 1.  
 XX  
 KW Human; dg; bisulphite treatment; CpG; DNA methylation; cancer; tumour;  
 KW cytoskeletal; ALDH6; CYP11A; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;  
 KW UDR8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200202806-A2.  
 XX  
 PD 10-JAN-2002.  
 XX  
 PF 29-JUN-2001; 2001WO-EP07470.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPig-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-154757/20.  
 XX  
 PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,  
 PT useful for detecting cytosine methylation state of genes associated  
 PT with pharmacogenomics and for therapy of diseases e.g. cancer -  
 XX  
 PS Claim 1; SEQ ID No 139; 24pp; English.  
 XX  
 CC The invention relates to a nucleic acid comprising a sequence at  
 CC least 18 bases in length of a segment of the chemically pretreated DNA  
 CC of genes associated with pharmacogenomics according to one of the  
 CC sequences of the genes ALDH6 (NM 000693), CYP11A (NM 000781), CYP11B1  
 CC (NM 000497), CYP3A3 (NM 000776 and NM 017460), DPYD (NM 000110), EPHX2  
 CC (NM 001979), OCLN (NM 002538), TXNRD1 (NM 003330), UDR8 (NM 003360),  
 CC MRP (NM 004996, NM 015900, NM 015901, NM 015902, NM 015962, NM 015988,  
 CC NM 015899) and their complementary sequences, or a sequence (SI) chosen  
 CC from 87 sequences and their complements. The chemical pretreatment  
 CC is bisulphite treatment to convert cytosines (but not methyl-cytosines)  
 CC into uracils. Also included are an oligomer (II) in particular an  
 CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in  
 CC each case at least one base sequence having a length of 9 nucleotides  
 CC which hybridises to or is identical to a chemically pretreated DNA of  
 CC genes associated with pharmacogenomics and their complements, arranged in  
 CC an array for analysing diseases associated with the methylation state  
 CC (pG) and/or detecting SNPs (single nucleotide polymorphisms)  
 CC of the 87 sequences. The oligomers may also be used as PCR primers.  
 CC The set of 87 nucleic acids and their complements is useful for diagnosis  
 CC and therapy of solid tumours and cancer. The present sequence  
 CC represents one the 87 DNA sequences or its complement.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 7441 BP; 2098 A; 66 C; 1948 G; 3329 T; 0 other;  
 XX  
 Query Match 1.1%; Score 20; DB 24; Length 7441;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1218 TTTTACACATATATAATA 1237  
 Db 6788 TTTTACACATATATAATA 6769  
 XX  
 RESULT 13  
 ID AAS63344/c  
 XX AAS63344 standard; DNA; 7479 BP.  
 AC AAS63344;  
 XX 29-JAN-2002 (first entry)

XX  
 DE Chemically pretreated metabolism associated gene #39.  
 XX  
 KW Human; cytostatic; anti-tumour; metabolism; metabolic disease; liver;  
 KW solid tumour; cancer; cytosine methylation; epigenetic; eye; kidney;  
 KW single nucleotide polymorphism detection; SNP; stool; urine; lung;  
 KW cerebral-spinal fluid; intestine; brain; heart; prostate; breast;  
 KW DUSP2; EPHX2; QDR; SGRH; SHMT2; SLC7A2; SLC7A4; TYMS; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200176451-A2.  
 XX  
 PD 18-OCT-2001.  
 XX  
 PF 06-APR-2001; 2001WO-EP04016.  
 XX  
 PR 06-APR-2000; 2000DE-1019058.  
 PR 07-APR-2000; 2000DE-1019173.  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPig-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-010834/01.  
 XX  
 PT New nucleic acid, useful for diagnosis and therapy of metabolic  
 PT disease, solid tumour and cancers, comprises segment of chemically  
 PT modified genomic sequences of genes associated with metabolism -  
 XX  
 PS Claim 1; Page 107-109; 143pp; English.  
 XX  
 CC The invention relates to a nucleic acid (II) comprising a sequence at  
 CC least 18 bases of a segment of the chemically pretreated DNA of genes  
 CC associated with metabolism such as DUSP2 (NM 004418), EPHX2 (NM 001979),  
 CC QDR (NM 000320), SGRH (NM 000199), SHMT2 (NM 005412), SLC7A2  
 CC (NM 003046), SLC7A4 (NM 004173) and TYMS (NM 001071) (all  
 CC undefined). (I) are useful for diagnosis and therapy of metabolic  
 CC disease, solid tumours and cancers; as primer oligonucleotides for the  
 CC amplification of DNA sequences; for detecting the cytosine methylation  
 CC state and/or single nucleotide polymorphisms (SNPs) in a chemically  
 CC treated DNA of genes associated with metabolism. An array of (I) is  
 CC useful for ascertaining genetic and/or epigenetic parameters for the  
 CC diagnosis and/or therapy of existing diseases or the predisposition to  
 CC specific diseases by analysing cytosine methylations. The method involves  
 CC chemically treating genomic DNA sample by a solution of bisulphite,  
 CC hydrogen sulphite or disulphite such that cytosine bases which are  
 CC unmodified at the 5th-position are converted to uracil or another base  
 CC which is dissimilar to cytosine in terms of hybridisation behaviour and  
 CC amplifying fragments of the chemically pretreated genomic DNA. The  
 CC genomic DNA is from cells or cellular components which contain DNA,  
 CC sources of DNA comprising, for e.g. cell lines, biopsies, blood, sputum,  
 CC stool, urine, cerebral-spinal fluid, tissue embedded in paraffin such as  
 CC tissue from eye, intestine, kidney, brain, heart, prostate, lung, breast  
 CC or liver, histologic object slides and their combinations. Genetic  
 CC parameters are mutations, in particular insertions, deletions, point  
 CC mutations, inversions and polymorphisms of genes associated with  
 CC metabolism and sequences further required for their regulation.  
 CC Epigenetic parameters are in particular cytosine methylations and  
 CC further chemical modifications of DNA bases of genes associated with  
 CC metabolism. Further epigenetic parameters include for e.g. the  
 CC acetylation of histones which correlates with DNA methylation.  
 CC AAS63306-AAS63373 represent chemically pretreated metabolism associated  
 CC genes, and related primers of the invention.  
 XX  
 SQ Sequence 7479 BP; 2111 A; 68 C; 1962 G; 3338 T; 0 other;  
 XX  
 Query Match 1.1%; Score 20; DB 24; Length 7479;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1218 TTTACACATATATAAATA 1237  
 |||||  
 DB 6826 TTTACACATATATAAATA 6807

RESULT 14  
 ID ABR80086 standard; DNA; 7922 BP.  
 AC ABR80086;  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Human chemically modified disease associated gene SEQ ID NO 103.  
 XX  
 KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
 KW antidiabetic; cyrostatic; anticonvulsant; ds.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO200200927-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07536.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIC-) EPICENOVICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130908/17.  
 XX  
 PT Novel nucleic acid useful for diagnosis and therapy of diseases  
 PT associated with development genes such as diabetes, comprises a  
 PT sequence of a segment of chemically pretreated DNA of genes associated  
 PT with development  
 XX  
 PS Claim 1; SEQ ID NO 103; 27bp; English.  
 XX  
 CC The invention relates to a nucleic acid (i) comprising a sequence at  
 CC least 18 bases in length of a segment of chemically pretreated DNA (ii)  
 CC of genes associated with development selected from 87 genes listed in  
 CC the specification such as ACCFN, ADPN, or AFDI and comprising one of 350  
 CC sequences (ABN79984-ABN80333) or their complements. The invention is  
 CC useful for the diagnosis or therapy of diseases associated with  
 CC development genes, in particular disease related to homeobox containing  
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
 CC associated with congenital heart disease, epilepsy, diseases related to  
 CC histone deacetylation, Curranio syndrome, diseases related with the  
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
 CC Oligomers specific to each of the genes are useful for detecting the  
 CC methylation state of all Cpg dinucleotides within the 350 sequences or  
 CC (ii) and their complementary sequences, as primer oligonucleotides for  
 CC the amplification of the 350 sequences, (ii) and/or their complements and  
 CC as oligomer probes for detecting the cytosine methylation state and/or  
 CC single nucleotide polymorphisms (SNPs).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office.  
 XX  
 SQ Sequence 7922 BP; 2094 A; 210 C; 1947 G; 3670 T; 1 other;

Query Match 1.1%; Score 20; DB 24; Length 7922;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY - 4 91g TTTAGTAGTATATAAGAA 937

DB 1568 TTTAGTAGTATATAAGAA 1587  
 |||||

RESULT 15  
 ID ABL29482 standard; DNA; 32548 BP.  
 AC ABL29482;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 39919.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical; gene; ds.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI; 2001-656860/75.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions  
 XX  
 PS Claim 1; SEQ ID NO 39919; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB116176-AB130511), expressed DNA  
 CC sequences (AB101840-AB116175) and the encoded proteins  
 CC (AB557737-AB12072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WPI  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 XX

Query Match 1.1%; Score 20; DB 23; Length 32548;  
 Best Local Similarity 100.0%; Pred. No. 43;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1108 GAAAAAATGTATATGAAT 1127  
 |||||  
 DB 30880 GAAAAAATGTATATGAAT 30899

Search completed: November 25, 2003, 17:02:58  
 Job time : 370 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 15:19:58 / Search time 4396 Seconds

(without alignments)  
16592.761 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atccactctctgactcttga.....tagagcgtacgaagctgaacc 1783

Scoring table:

OLIGO\_NTC  
Gapop 60.0, Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

GenEmbl: \*  
1: gb\_ba: \*  
2: gb\_hlg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_ov: \*  
22: em\_ox: \*  
23: em\_pac: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_hlg\_hum: \*  
31: em\_hlg\_inv: \*  
32: em\_hlg\_other: \*  
33: em\_hlg\_mus: \*  
34: em\_hlg\_pln: \*  
35: em\_hlg\_rdc: \*  
36: em\_hlg\_mam: \*  
37: em\_hlg\_vrt: \*  
38: em\_sy: \*  
39: em\_hlgo\_hum: \*  
40: em\_hlgo\_mus: \*  
41: em\_hlgo\_other: \*

score greater than or equal to the score being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	24	1.3	180097	9	AC123978 Papio anu
2	24	1.3	227691	2	AC110885 Mus muscu
3	23	1.3	46067	10	AF539527 Mus muscu
4	23	1.3	120648	8	AF003859 Oryza sat
5	23	1.3	130689	2	AL806534 Mouse DNA
6	23	1.3	169888	2	AC115848 Mus muscu
7	23	1.3	303446	1	AB016941 Bacteroid
8	22	1.2	2282	8	VNA277287 Vicia fab
9	22	1.2	67271	2	AC133887 Homo sapi
10	22	1.2	108653	9	HSJD863C7 Human DNA
11	22	1.2	153537	2	AC121962 Mus muscu
12	22	1.2	160536	2	AC040953 Homo sapi
13	22	1.2	165870	9	AC025089 Homo sapi
14	22	1.2	175667	9	AC009336 Homo sapi
15	22	1.2	194149	9	AC100799 Homo sapi
16	22	1.2	201923	10	AC121772 Mus muscu
17	22	1.2	201935	10	AC121582 Mus muscu
18	22	1.2	206975	2	AC133827 Rattus no
19	22	1.2	212024	2	AC073390 Mus muscu
20	22	1.2	224846	10	AC125327 Mus muscu
21	22	1.2	236579	2	AC121707 Rattus no
22	22	1.2	245226	2	AC103003 Rattus no
23	22	1.2	290670	2	AC121043 Rattus no
24	21	1.2	8011	6	AX344981 Sequence
25	21	1.2	8011	6	AX344981 Sequence
26	21	1.2	9060	1	U39721 Mycoplasma
27	21	1.2	11907	6	AX344316 Sequence
28	21	1.2	19366	7	PZACG M1813 Bacterioph
29	21	1.2	36087	3	L16685 Caenorhabdi
30	21	1.2	42732	9	AC007206 Homo sapi
31	21	1.2	43157	9	AC079466 Homo sapi
32	21	1.2	80059	9	AC114940 Homo sapi
33	21	1.2	90500	9	AC069235 Homo sapi
34	21	1.2	100926	9	AF583802 Human DNA
35	21	1.2	108537	2	AC139355 Medicago
36	21	1.2	110000	3	AC116957 Dictyoste
37	21	1.2	110000	6	Continuaction (2 of
38	21	1.2	110000	6	Continuaction (5 of
39	21	1.2	125400	9	AC123904 Homo sapi
40	21	1.2	129451	9	AL359474 Human DNA
41	21	1.2	134904	9	AC004879 Homo sapi
42	21	1.2	137783	9	AL441884 Human DNA
43	21	1.2	139378	9	AL031668 Human DNA
44	21	1.2	142805	9	AL592525 Human DNA
45	21	1.2	147038	9	AL133326 Human DNA

#### ALIGNMENTS

RESULT 1  
AC123978 180097 bp DNA linear PRI 01-AUG-2002  
LOCUS Papio anubis clone rp41-211ds, complete sequence.  
AC123978  
ACCESSION AC123978  
VERSION AC123978.5 GI:22038600  
KEYWORDS HTG.  
SOURCE Papio anubis (olive baboon)  
ORGANISM Papio anubis  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
Cercopithecinae; Papio.  
REFERENCE  
1 (bases 1 to 180097)  
Song, L. and Roe, B.A.  
Papio anubis BAC Clone rp41-211ds

Pred. No. is the number of results predicted by chance to have a

JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 180097)  
 AUTHORS Song, L. and Roe, B. A.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-JUN-2002) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

REFERENCE 3 (bases 1 to 180097)  
 AUTHORS Song, L. and Roe, B. A.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-JUL-2002) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

REFERENCE 4 (bases 1 to 180097)  
 AUTHORS Song, L. and Roe, B. A.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-AUG-2002) Department Of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA

COMMENT -- Genome Center  
 -- Center: Department Of Chemistry And Biochemistry  
 -- The University Of Oklahoma  
 -- Center code: UOKNOR

FEATURES  
 Source  
 1..180097  
 /organism="Papio anubis"  
 /mol\_type="genomic DNA"  
 /sub\_species="anubis"  
 /db\_xref="taxon:9555"  
 /clone="rp41-211d5"  
 /clone\_lib="RPCI - 41 Male (Olive) Baboon BAC Library"

BASE COUNT 48904 a 39123 c 39387 g 52683 t

ORIGIN

Query Match 1.3%; Score 24; DB 9; Length 180097;  
 Best Local Similarity 100.0%; Pred. No. 1.2;  
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 314 ATTCTTACACGTGAGAAATTT 397  
 |||||  
 Db 115577 ATTCTTACACGTGAGAAATTT 115600  
 |||||

RESULT 2  
 LOCUS AC110885 227691 bp DNA linear HTG 21-AUG-2002  
 DEFINITION Mus musculus clone RP24-342P18, WORKING DRAFT SEQUENCE, 9 ordered  
 pieces.  
 AC110885  
 AC110885.5 GI:22381716  
 HTG: HTGS PHASE2; HTGS DRAFT; HTGS\_FULLTOP.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 227691)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP24-342P18  
 Unpublished  
 2 (bases 1 to 227691)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckigalter, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Charao, B.,  
 Chesepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Farc, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Girard, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kanat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R.,  
 Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,  
 MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,

JOURNAL Unpublished  
 REFERENCE 3 (bases 1 to 227691)  
 AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,  
 Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckigalter, B.,  
 Camarata, J., Chang, J., Charao, B., Chesepel, Y., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
 Farc, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,  
 Gardyna, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B.,  
 Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A.,  
 Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,  
 Liu, G., Maclean, C., MacDonald, P., Major, J., Matthews, C.,  
 McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severy, P.,  
 Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talmas, J.,  
 Telford, S., Theodore, J., Topham, K., Travers, M., Vassiliou, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
 Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Aug 21, 2002 this sequence version replaced gi:20506336.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIRB  
 Web site: http://www-seq.wi.mit.edu  
 Contact: sequence\_submissions@genome.wi.mit.edu  
 ----- Project Information  
 Center project name: 119395  
 Center clone name: 342.P.18  
 ----- Summary Statistics  
 Sequencing vector: plasmid, N/A; 100% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 225012 bases at least Q40  
 Consensus quality: 226175 bases at least Q30  
 Consensus quality: 226632 bases at least Q20  
 Insert size: 22500; agarose-fp  
 Insert size: 226891; sum-of-contigs  
 Quality coverage: 12.1 in Q20 bases; agarose-fp  
 Quality coverage: 12.0 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently  
 consists of 9 contigs. Gaps between the contigs  
 are represented as runs of N. The order of the pieces  
 is believed to be correct as given, however the sizes  
 of the gaps between them are based on estimates that have  
 been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.  
 1 2355: contig of 2355 bp in length  
 \* 2356 2455: gap of 100 bp  
 \* 2456 10338: contig of 7883 bp in length  
 \* 10339 10438: gap of 100 bp



FEATURES	source	1. .227691	/organism="Mus musculus"	
			/mol_type="genomic DNA"	
			/dd_xref="taxon:10090"	
			/clone="RP24-342P18"	
			/clone_11b="RPC1-24 Male Mouse BAC"	
		1. .2355		
misc_feature		/note="assembly_fragment"		
misc_feature		2456. 10338		
misc_feature		/note="assembly_fragment"		
misc_feature		10439. 17967		
misc_feature		/note="assembly_fragment"		
misc_feature		18068. 27410		
misc_feature		/note="assembly_fragment"		
misc_feature		27511. 44269		
misc_feature		/note="assembly_fragment"		
misc_feature		44370. 65178		
misc_feature		/note="assembly_fragment"		
misc_feature		65279. 89146		
misc_feature		/note="assembly_fragment"		
misc_feature		89247. 137077		
misc_feature		/note="assembly_fragment"		
misc_feature		137178. .227691		
		/note="assembly_fragment"		
		clone_end:"77"		
		vector_side:"right"		
BASE COUNT		61965 a 50768 c 49660 g 64498 t		800 others
ORIGIN				
Query Match		1.3%; Score 24; DB 2; Length 227691;		
Best Local Similarity		100.0%; Pred. No. 1.1;		
Matches		24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	972	TTTAAATTTGATTAAGTTGTAT 995		
Db	161120	TTTAAATTTGATTAAGTTGTAT 161143		
RESULT 3				
AF539527				
LOCUS				
DEFINITION		Mus musculus strain C57BL/6 O-linked GlcNAc transferase gene,		
ACCESSION		complete cds.		
VERSION		AF539527		
KEYWORDS		AF539527.1 GI:27499605		
SOURCE				
ORGANISM		Mus musculus (house mouse)		
		Mus musculus		
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
		1 (bases 1 to 46067)		
REFERENCE		Hanover,J.A., Yu,S., Lubas,W.B., Shin,S.H., Ragano-Caracciola,M.,		
AUTHORS		Kochian,J. and Love,D.C.		
TITLE		Mitochondrial and nucleocytoplasmic isoforms of O-linked GlcNAc		
		transferase encoded by a single mammalian gene		
		Arch. Biochem. Biophys. 403 (2), 287-297 (2003)		
JOURNAL		2 (bases 1 to 46067)		
PUBMED		12504895		
REFERENCE		Hanover,J.A., Yu,S., Lubas,W.B., Shin,S.-H., Ragano-Caracciola,M.,		

TITLE J.K. and Loye,D.C.  
 JOURNAL Direct Submission  
 Submitted (09-AUG-2002) LCB, NIDK, NIH, Bldg 8 Rm 402, 9000  
 Rockville Pike, Bethesda, MD 20892-0851, USA  
 FEATURES Location/Qualifiers  
 source  
 1. .46067  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /strain="C57BL/6"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 /map="XD"  
 mRNA  
 join(1309. .1554,4215. .4395,5188. .5311,13392. .13460.  
 16877. .16993,22505. .22584,23623. .23818,24590. .24730.  
 25393. .25493,28017. .28170,28255. .28356,28719. .28858.  
 30500. .30658,30774. .30863,33191. .33316,33786. .33956.  
 34066. .34182,34285. .34455,36170. .36322,38732. .38954.  
 39247. .39370,41908. .44103)  
 /product="O-linked G1CNAC transferase"  
 join(1518. .1554,4215. .4395,5188. .5311,13392. .13460.  
 16877. .16993,22505. .22584,23623. .23818,24590. .24730.  
 25393. .25493,28017. .28170,28255. .28356,28719. .28858.  
 30500. .30658,30774. .30863,33191. .33316,33786. .33956.  
 34066. .34182,34285. .34455,36170. .36322,38732. .38954.  
 39247. .39370,41908. .42082)  
 /codon\_start=1  
 /product="O-linked G1CNAC transferase"  
 /protein\_id="AA017363.1"  
 /db\_xref="GI:27495606"  
 /translation="MASSGVNADSTPTPTRMISFOGLAEIHHREYQGFPAERH  
 MQMRDEPNTGALLILISHPQCRIDRSAPHESTLAKNPILAEEVSNIGNVKEE  
 GQGLAEIHHREYHALLRKPDIDIGYINAAALVAGDVEGVAQVAVSLAQNPDYCYKE  
 SDGLNALKALGRLEBAKCYLKALETQPNFPAVAGSNIGVNAAGSEILAIHREKAL  
 TLDPNFIADYINIGNVLEKARIIPRAVAAYLRALSIPNHAIVGNGLCYVEGGIL  
 LAIDYRRATEILOHPFDPAKCAIRIPDAVAALCNLANIKEGSVAEARDCYNTALRCP  
 NIKREQENITEARVLRKALEVEPPEFAASNLASVYIQOQCKLOEALMRYKPAIRIS  
 TFDAYVNSGNTLTKEMQDVQALQCYTRAIQINFAFADASHNLASIHDSNITEAL  
 SYSTALILKPDFPDPAKCAIRIPDAVAALCNLANIKEGSVAEARDCYNTALRCP  
 HSMYLIHSGFRKAIARHGNLCIDKINVLAKPRYEPKDLKSDGLRIVGVSDFG  
 NHFSLPMQSIPOKEMNDKFEVXCYSALPDQGNFPRKVAEAAHFTLSQIPCNQK  
 ADRHQGHGHIIVNNGTYTGARENELFALPAPLOAMWIGPNSGALPMYITLDDQ  
 TSPAEVDEQYSEKLIAYMPPHPEFIDGNAHNPBKAKKVIPEKSGHLYNDNIVNGIT  
 LKAFDLSLDPVKTYVMKCPDGGNDPDSNTALMNPVTPMNTIARAVTEIMNRGQIQIT  
 INGSISNGIATQIINNKAATGEEVEPTIIIVTYSQGLPEDAIVNFNQLYKIDP  
 TLQMWANILKRVNSVYMLMRFPVAGVEPITQVYQNNGLDPNRIIFPVARKEDHVR  
 GQIADVICLIDPLCNGHTGMDVMAGCPMTMPEETILASRAAQILCLGCEILAKRS  
 ROEEDIAVLKIDGLDELYELKKIRGVKQRKQRISSPLFNXXQYTMELERLYLQWHEHYAAG  
 NKDDHMKPEVTESA"

BASE COUNT 12287 a 9079 c 9503 g 15155 t 43 others

ORIGIN

Query Match 1.3%; Score 23; DB 10; Length 46067;  
 Best Local Similarity 100.0%; Pred. No. 4.5;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 430 AAACCTTGACTTTTGACTTTCTTA 452  
 |||||  
 Db 22853 AAACCTTGACTTTTGACTTTCTTA 22875

RESULT 4  
 AP003859/c  
 LOCUS AP003859 120648 bp DNA linear PLN 12-JUN-2003  
 DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 8,  
 BAC clone:OJ1033\_B09, complete sequence.  
 ACCESSION AP003859  
 VERSION AP003859.2 GI:31621035  
 KEYWORDS HTG.  
 SOURCE Oryza sativa (japonica cultivar-group)  
 ORGANISM Oryza sativa (japonica cultivar-group)  
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatopsida; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 Ehrhartoideae; Oryzaceae; Oryza.

REFERENCE 1  
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.  
 TITLE *Oryza sativa nipponbare* (GA3) genomic DNA, chromosome 8, BAC clone:OJ1033 B09  
 JOURNAL Published Only in Database (2001)  
 REFERENCE 2 (bases 1 to 120648)  
 AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JUL-2001) Takuji Sasaki, National Institute of Agricultural Sciences, Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan  
 E-mail: tsasaki@nias.affrc.go.jp, URL: http://rsg.dna.affrc.go.jp/, Tel: 81-298-38-7441, Fax: 81-298-38-7468  
 COMMENT On Jun 12, 2003 this sequence version replaced gi:14646792. The orientation of the sequence is from M13rev to -21M13 of the BAC clone. The nucleotide sequence of this BAC clone was generated by combining Monsanto and RGP-Japan sequencing data.

FEATURES  
 source  
 Location/Qualifiers  
 1..120648  
 /organism="Oryza sativa (japonica cultivar-group)"  
 /mol\_type="genomic DNA"  
 /cultivar="Nipponbare"  
 /db\_xref="taxon:39947"  
 /chromosome="8"  
 /clone="OJ1033\_B09"

BASE COUNT 35274 a 25678 c 25266 g 34430 t  
 ORIGIN

Query Match 1.3%; Score 23; DB 8; Length 120648;  
 Best Local Similarity 100.0%; Pred. No. 3.9;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1363 TTATTTATATTGTGCATATT 1385  
 |||||  
 Db 88378 TTATTTATATTGTGCATATT 88356

RESULT 5  
 LOCUS AL806534 130689 bp DNA linear ROD 19-SEP-2002  
 DEFINITION Mouse DNA sequence from clone RP23-334A11 on chromosome X, complete sequence.  
 ACCESSION AL806534 GI:23304104  
 VERSION AL806534.3  
 KEYWORDS HTG.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 130689)  
 Blakey, S.  
 Direct Submission  
 Submitted (17-SEP-2002) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 On Sep 23, 2002 this sequence version replaced gi:22798318.

COMMENT  
 -----  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: http://www.sanger.ac.uk  
 Contact: humquerry@sanger.ac.uk  
 -----  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; SW, SwissProt; Tr, TrEMBL; Wp, WormPep; information on the NORMPep database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormPep RP23-334A11 is from the RPI-23 Mouse PAC Library  
 constructed by the group of Pieter de Jong.  
 For further details see http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6  
 Location/Qualifiers  
 1..130689  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 /clone="RP23-334A11"  
 /clone\_id="RPI-23"

BASE COUNT 37078 a 25556 c 27928 g 40127 t  
 ORIGIN

Query Match 1.3%; Score 23; DB 10; Length 130689;  
 Best Local Similarity 100.0%; Pred. No. 3.9;  
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 430 AAACCTTGACCTTTGACCTTCTTA 452  
 |||||  
 Db 119793 AAACCTTGACCTTTGACCTTCTTA 119815

RESULT 6  
 LOCUS AC115848 169888 bp DNA linear HTG 23-APR-2003  
 DEFINITION Mus musculus clone RP24-248L23, WORKING DRAFT SEQUENCE, 12 ordered pieces.  
 ACCESSION AC115848 GI:30017915  
 VERSION AC115848.3  
 KEYWORDS HTG, HTGS\_PHASE2, HTGS\_DRAFT.  
 SOURCE Mus musculus (house mouse)  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 169888)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP24-248L23  
 Unpublished  
 2 (bases 1 to 169888)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chararo, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A., Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fato, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyne, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, M., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kelle, C., Lacombe, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Mercus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicot, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhan, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Theodore, J., Tophan, K., Travers, M., Travis, N., Triggilo, J., Vasiliiev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCES  
AUTHORS

3 (bases 1 to 16988)

Birren, B., Nusbaum, C., Lander, E., Abouelella, A., Allen, N., Anderson, S., Arachchi, H.M., Barna, N., Bastien, Y., Bloom, T., Engelslavsky, L., Soukhtgatter, B., Camarata, J., Chang, Y., Choepel, Y., Collumore, A., Cook, A., Cooke, P., Corum, B., Deaveland, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erikson, T., Fard, S., Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Haez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamet, A., Kartas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Munga, V., Murphy, T., Naylor, J., O'Neill, D., Oliver, J., Peterson, K., Plunkhard, P., Pierre, N., Rachubka, A., Rameray, U., Raymond, C., Reiter, R., Rise, C., Roman, U., Schauer, S., Schnupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M., Talanas, J., Teflate, S., Theodocje, J., Topham, K., Travers, M., Vassiliou, H., Venkataratnam, V.S., Viet, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission

Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 17, 2003 this sequence version replaced gi:21431360.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

```

/db xref="taxon:10090"
/clone="RP24-248L23"
/clone_lib="RPCT-24 Male Mouse BAC"
misc_feature
1..1143
/note="assembly_fragment"
1244..12026
/note="assembly_fragment"
2127..3200
/note="assembly_fragment"
3301..4372
/note="assembly_fragment"
4473..5884
/note="assembly_fragment"
5985..7872
/note="assembly_fragment"
7973..10011
/note="assembly_fragment"
10112..17325
/note="assembly_fragment"
17426..40917
/note="assembly_fragment"
50018..97580
/note="assembly_fragment"
97681..148528
/note="assembly_fragment"
148629..169868
/note="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT      50714 a 33304 c 33190 g 51576 t 1104 others
ORIGIN
Query Match      1.3%; Score 23; DB 2; Length 169888;
Best Local Similarity 100.0%; Pred. No. 3.7;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0.
OY      1031 GATTTTGAAATTATGTTTCTGA 1053
|||||
|||
Db      28640 GATTTTGAAATTATGTTTCTGA 28662

```

gene  
CDS  
/locus\_tag="B73514"  
178..3126  
/locus\_tag="B73514"  
/codon\_start=1  
/transl\_table=11  
/product="conserved hypothetical protein"  
/protein\_id="AA078620.1"  
/db\_xref="GI:29340830"  
/translation="MRILPLVLSTSLAQSQVIREKNAIQSKAFPLVTSKAAV  
VYDIDYLVKRTAKLEFSDVESVTGQQLTDKLGDKELIIVGVEKNRLIRLQAE  
KGLDLSLEGTEWELFIQVSRPFGVSVALVAGSDRGAAVGLFISEMNGSPW  
VMADYVKGKKTLYVADPATLSKTSVSKRGIFLNDMDGLKPMNAKTFEKEKRGIG  
PRVYAKICELLRLKANHLLAPAMPVSTAFYKIPENKLVADTFAIMGSHCEPLIN  
TASBMSKTMGPMDYNKNDKINENLSNRYKCAVENEYATLRLGLHAAAGGSDVP  
MRKVKMLBSALAKREILRRHIDKVEVETIPQAFYKYVLEIYSGLLEPDDVTIN  
ADNFGMKRLSGPQKSGRAGVYHISLSPASPHSYMTITTPALMKEEYVDT  
TTADRWLANCGDLKGAETQISLFLMAYDISEFNADNVAIYPARMLAKFEEYVDT  
LEDITCSHLNLFASRKEVYMGWYNNYGGGKRTDTEFSFANYEARRLTEYSRI  
GKKTEDLASLDEKSKPAFYQLYEVKGLMANNMTIKGQFYRQYVQRAANRLK  
AAYKCHDSLEITTDGYNLLDGKMHMSLKONYDGTSSYEMIPLMEBYTPVGPBK  
LGIQABSENIDKSGMSPHTLPASTYSRKSHDIDYOGTGLMSITPSEDMTILISQ  
KSGKTSANRITYISVBDKYPAGEKRGQDIDISGQKSTVYVSNPSPSPARTVQ  
LYVEENGYISIPADPHRKESNDIMMSILPGLFGRSLQGNPTAPLQYRAGDVP  
RYEYPTFNAGYDYVTVLPTPLAERDYKLPEHTNSDTKYSVRIDDSISTPST  
SALEYQIWDVSLKNCRVKSTLYVKKPGKHTLQRCDDPGVVIQKIYIDGKMS  
YLGDSITICN"  
complement (3123..4292)  
/locus\_tag="B73515"  
complement (3123..4292)  
/locus\_tag="B73515"  
/codon\_start=1  
/transl\_table=11  
/product="putative glycosylhydrolase"  
/protein\_id="AA078621.1"  
/db\_xref="GI:29340831"  
/trans\_action="W1QYFYLHGHYQMSRMSRLPLSLIILIGVCOHKRAT  
EKETVAGNTYHPLERGAEPWALFHEGKTYTQSGENKQILMETDDITDLSHATQ  
DWIPLDPSNHYMLPEIRINNKKWYIYPAADGMNDNQIIVENENAMPEGTV  
MKGIQTDKNNMHAISTEHDQRYMIWCGQKRIIDETQCIYAMKRPWLTSS  
DRILSKREYEMECOWPDGSKTAYIHNVEAPYFESKNCKACIFYSAGSGMTPY  
YCVGLPADAKANLNPASWKSFPVLPQDDPENNYPGGISTPSPGKEVMYLYH  
ARQIPNDAPGASDSRRLQKIDMDQGMVLTGPKEEPMARPSGSPHKKNQ"  
4402..6315  
/locus\_tag="B73516"  
4402..6315  
/locus\_tag="B73516"  
/codon\_start=1  
/transl\_table=11  
/product="arabinan endo-1,5-alpha-L-arabinosidase A  
precursor"  
/protein\_id="AA078622.1"  
/db\_xref="GI:29340832"  
/translation="MKLRHLFLFCVLCQSVSISAONKSKLPKTSGNPIFPGVADBE  
IVGDEWYIPTYSAAYDQIFMDAFSSKDLVMTKHPKYLSENISMLRALMAVAV  
IHANDKYFFFGANDIONNNELGIGVAVADNPAGPKALGKPLIDKIVNGAQPIDQ  
FVKKDDGQYVNYGGMGHCNMVAPADLLSTVPEEDGTYKEVTPONYVEGPMXK  
NGKTYFMSBGGGDPYVAIYLAUSPPRPFERVKILQODTNINATGAGHSVHVC  
GSDWYLIYHRRLPGETAANSRATCIDRMFFNKDGLFVRMTFEGVAGSPLTPRA  
KTYANPYNFSLDPTIIGXDEVEYLYATESIKNPIHRSRLVWYVYGTAFYAT  
RPNPEKGRIVADPINKIKGKYVWYMSWTGEMTCIGIATADKREGFPTDGLKLF  
RSNEINONCIDPFIYEDNGKLYFMGSHGIGYELTDLGLLKKAPAEQIAGTAY  
EGTYIHRKDGWYFASIGRCCBGLKSTYTTVGRSKNLPFYVDDKKGRMSNHHEI  
LIQCKKAFVGGHSELVTDKKNDMFHYAVSVANEGEVLMDNRYQMGMPFYVKG  
AVPSLETAADPDR"  
6571..7143  
/locus\_tag="B73517"  
6571..7143  
/locus\_tag="B73517"  
/codon\_start=1  
/transl\_table=11  
/product="RNA polymerase Ecf-type sigma factor"  
/protein\_id="AA078623.1"  
/db\_xref="GI:29340833"

gene  
CDS

/translation="MASDHTLSDLSLMKRFLEGGSSAQIYNQVQLFRYGLVY  
SDRELVDCTHDVFKIYTRAKLTFDNIIVAIMVALKNTLFPALKTSDSFSLDA  
DKEDQSEERSTPETIYINKDEKXNTHVMMSLSLTTRQREIYVRYIKMSIDE  
ISKITDNNYQSVNSISORALGRVNLFKRE"  
7257..8459  
/locus\_tag="B73518"  
7257..8459  
/locus\_tag="B73518"  
/note="Tirf homolog"  
/codon\_start=1  
/transl\_table=11  
/product="putative anti-sigma factor"  
/protein\_id="AA078624.1"  
/db\_xref="GI:29340834"

gene  
CDS

/translation="MINSRLKHTDFSRFTFEEFLQNDFFISVYKPTBEIOEFMRPE  
KSTPSINDYVPAAREYLETJSTSGDLSLOEGLMADIQTNIKDKIKHNFLLI  
GLTAAVAILVGSFELLKNVOALADINTFAVQTELPATEETLIILAEDEVVL  
KEKTTTIVDSVALKABENISKEALANVNOIYPRKRSIVLPPDSQSKVAVAGTV  
IYTFEPEDKREIYVDEIITIVARDERPFYRTTRMVRVLQTKNNYAVASEPR  
SVLVAQGVQVETTQTKAILAPQMSVBEKENISQVDEQMTISVNGLYCPNSAD  
LGIYIKRLSTYYGINVEFDSLSKIKSGKIDLKDNFETVINGLTFVAFISVAYDQY  
KTYAVKK"  
8596..12051  
/locus\_tag="B73519"  
8596..12051  
/locus\_tag="B73519"  
/note="SusC homolog"  
/codon\_start=1  
/transl\_table=11  
/product="putative outer membrane protein, probably  
involved in nutrient binding"  
/protein\_id="AA078625.1"  
/db\_xref="GI:29340835"

gene  
CDS

/translation="MKKRYNLLGLSLFVTTLLSGDYFDQVDDRLSLEELFKTR  
DQALYLSNVYFLPDEPNQROVHTSLYRIPGWPAGSDTEMWTSNGKAKILNNS  
IDATNTWLVLRKMSWISGHSRAFTTYVQAPLTAERAOQWAEALAIAYEVL  
VRTGVPVLEDDVALDTPSNELQSRSTVRCDFIVSELKEAQNALLEDASSDKT  
TGVGRIDBAIQAFIIEALTYRASMLFNGECTIYVADMANPQGRILPESQPDATIKAD  
KOKVTECOKFAADYGNFOLMYTDDKSGSVAGBPADEFNYESYAIRTLFSEMG  
NKENI FYMDNAGTMOYDRMKNKSGNTNDYRGSLIGATQTEWYVADYFAMGMSVTC  
VADGVTPILINASGYKDKDSSSDYSKADGLVAPACTRNMVYVREBRFVADITFSN  
VNPF"  
12064..13992  
/locus\_tag="B73520"  
12064..13992  
/locus\_tag="B73520"  
/note="Susd homolog"  
/codon\_start=1  
/transl\_table=11  
/product="putative outer membrane protein, probably  
involved in nutrient binding"  
/protein\_id="AA078626.1"  
/db\_xref="GI:29340836"

/translation="MKKRYNLLGLSLFVTTLLSGDYFDQVDDRLSLEELFKTR  
DQALYLSNVYFLPDEPNQROVHTSLYRIPGWPAGSDTEMWTSNGKAKILNNS  
IDATNTWLVLRKMSWISGHSRAFTTYVQAPLTAERAOQWAEALAIAYEVL  
VRTGVPVLEDDVALDTPSNELQSRSTVRCDFIVSELKEAQNALLEDASSDKT  
TGVGRIDBAIQAFIIEALTYRASMLFNGECTIYVADMANPQGRILPESQPDATIKAD  
KOKVTECOKFAADYGNFOLMYTDDKSGSVAGBPADEFNYESYAIRTLFSEMG  
NKENI FYMDNAGTMOYDRMKNKSGNTNDYRGSLIGATQTEWYVADYFAMGMSVTC  
VADGVTPILINASGYKDKDSSSDYSKADGLVAPACTRNMVYVREBRFVADITFSN  
VNPF"

```

SKWPGTEGDTYVDFYSGNGKAGQANNDFSTGYIVKXNMGSDPRNOLVCVLLBLT
NIYFDYIEALAYVNDPSHADTKMKNLLPERAGIGCYGTATSLPRATTTSEMLKIQEK
RIELSEFNCXKFDYRWGLTNEFFNFKVHGNNVYDGNFFYKTEILTSRNFDRQFFP
IPGSEIDIDKXLVQNEGF"
14018..15307
/locus_tag="BT3521"
14018..15307
/locus_tag="BT3521"
/codon_start=1
/trans_table=11
/product="alpha-1,6-mannanase"
/protein_id="AA078627.1"
/db_xref="GI:29340837"
/translation="MMKQYIFSAICLVSAFCCLSSGNDCEARPYTPDYRIVEEYTN
DTWKAYAEFNEHLIDONKPIYKSTADKAAVDWNGAAATWOCPTWDMANAKYAK
AEQPTCKEKKFKQ.CDLPFAGNKAHYNPFDDNNENTGPFITDDIKWTVTLARAYE
LRGYEYILSLSEBSFRWYSGEKVGTGSADPEKLGSGKMWMQPIKNPNENAD
HGKACINPFTVAALTLVNNVPGTESIDHSYQTKSQYAKGKEIYAMVENLV
DVTGGVADSRHNGNPNAMKDHVKNQSYGASVLYKATGEKQYLDNAMADYTN
TISGTPDLPEFTGAEQIYTAFAQYIAMLVYDCDQTOYIPVKENINVMANREDKT
RDICGGDYTKLQYGDVASEYSASGIPALMLPPTDK"
15321..16493
/locus_tag="BT3522"
15321..16493
CDS
gene
Query Match 1.3%; Score 23; DB 1; Length 303446;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 411 ATTTGACTTTCCAAAGCAAC 433
Db 178526 ATTTGACTTTCCAAAGCAAC 178504

RESULT 8
VFA277287/c
LOCUS VFA277287 2282 bp DNA linear PLN 07-DEC-2000
DEFINITION Vicia faba partial enod18 gene, exon 1 and 5' UTR.
ACCESSION A277287.1 GI:11602752
VERSION A277287.1 GI:11602752
KEYWORDS enod18 gene.
SOURCE Vicia faba (fava bean)
ORGANISM Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae;
Vicia.
1
REFERENCE
1 Hohnjec,N., Kuster,H., Albus,U., Frosch,S.C., Becker,J.D.,
Puhler,A., Perlick,A.M. and Fruhling,M.
TITLE The broad bean nodulin VFNOD18 is a member of a novel family of
plant proteins with homologies to the bacterial Mj0577 superfamily
JOURNAL Mol. Gen. Genet. 264 (3), 241-250 (2000)
MEDLINE 20535713
PubMed 11085263
2 (bases 1 to 2282)
REFERENCE
1 Kuester,H.
TITLE Direct Submission
JOURNAL Submitted (05-APR-2000) Kuester H., Leibnizuni fuer Genetik,
Universitaet Bielefeld, Postfach 100131, NRW, GERMANY
FEATURES
source
1..2282
/organism="Vicia faba"
/mol_type="genomic DNA"
/strain="Kleine Thueringer"
/db_xref="taxon:3906"
/tissue_lib="nodule cDNA library"
1948..2213
/gene="enod18"
1948..1958
/gene="enod18"
1957..2213
/gene="enod18"
1957..2213
/gene="enod18"

```

```

1997..2213
/gene="enod18"
/number=1
2039..2213
/gene="enod18"
/codon_start=1
/product="ENOD18 protein"
/protein_id="CAC18559.1"
/db_xref="GI:11602753"
/translation="MVEDRKVGVGIDFSKSNALKMAIVNADKDFEYLHINSNS
SDERSKLFKAKTS"
BASE COUNT 772 a 376 c 329 g 805 t
ORIGIN
Query Match 1.2%; Score 22; DB 8; Length 2282;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 887 TTAATTAATGATGTTGTAAC 908
Db 1271 TTAATTAATGATGTTGTAAC 1250

RESULT 9
AC133887
LOCUS AC133887 67271 bp DNA linear HTG 19-SEP-2002
DEFINITION Homo sapiens clone CTD-2584F10, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC133887.1 GI:23196488
VERSION AC133887.1 GI:23196488
KEYWORDS HTG; HTGS_PHASEO.
SOURCE Homo sapiens (human)
ORGANISM Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 67271)
REFERENCE
1 Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone CTD-2584F10
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 67271)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boucknight,B.,
Camarata,O., Chang,J., Chazaro,B., Choepel,Y., Collumore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrum,J., Meneses,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,D., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupack,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Strange-Thomann,N., Stojanovic,N., Talamas,J.,
Testaye,S., Theodore,J., Topfman,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (19-SEP-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
Project Information
Center Project name: L28217
Center clone name: 2584_F_10
* NOTE: This record contains 81 individual

```

\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.

1  
716  
816  
1549: contig of 715 bp in length  
816  
1549: contig of 734 bp in length  
1550  
1649: gap of 100 bp  
2386: contig of 737 bp in length  
1550  
2386: contig of 737 bp in length  
2387  
2486: gap of 100 bp  
2487  
3215: contig of 729 bp in length  
2487  
3215: contig of 729 bp in length  
3216  
4036: gap of 100 bp  
3216  
4036: contig of 721 bp in length  
4037  
4136: gap of 100 bp  
4037  
4875: contig of 739 bp in length  
4876  
4975: gap of 100 bp  
4876  
5715: contig of 740 bp in length  
5716  
5815: gap of 100 bp  
5816  
6540: contig of 725 bp in length  
6541  
6640: gap of 100 bp  
6641  
7375: contig of 735 bp in length  
7376  
7475: gap of 100 bp  
8215: contig of 740 bp in length  
8216  
8315: gap of 100 bp  
8316  
9053: contig of 738 bp in length  
9054  
9153: gap of 100 bp  
9154  
9889: contig of 736 bp in length  
9890  
9889: gap of 100 bp  
9890  
10729: contig of 740 bp in length  
10730  
10829: gap of 100 bp  
10830  
11558: contig of 729 bp in length  
11559  
11658: gap of 100 bp  
11659  
12390: contig of 732 bp in length  
12391  
12490: gap of 100 bp  
12491  
13235: contig of 745 bp in length  
13236  
13335: gap of 100 bp  
13336  
14066: contig of 731 bp in length  
14067  
14166: gap of 100 bp  
14167  
14894: contig of 728 bp in length  
14895  
14994: gap of 100 bp  
14895  
15718: contig of 724 bp in length  
15719  
15818: gap of 100 bp  
15819  
16553: contig of 735 bp in length  
16554  
16653: gap of 100 bp  
16654  
17387: contig of 734 bp in length  
17388  
17487: gap of 100 bp  
17488  
18229: contig of 742 bp in length  
18229  
18329: gap of 100 bp  
18230  
18329: gap of 100 bp  
18330  
19062: contig of 733 bp in length  
19063  
19162: gap of 100 bp  
19063  
19892: contig of 730 bp in length  
19893  
19992: gap of 100 bp  
19893  
20739: contig of 747 bp in length  
20739  
20839: gap of 100 bp  
20840  
21579: contig of 740 bp in length  
21580  
21679: gap of 100 bp  
21680  
22426: contig of 747 bp in length  
22427  
22526: gap of 100 bp  
22527  
23265: contig of 739 bp in length  
23266  
23365: gap of 100 bp  
23366  
24100: contig of 735 bp in length  
24101  
24200: gap of 100 bp  
24201  
24944: contig of 744 bp in length  
24945  
25044: gap of 100 bp  
25045  
25776: contig of 732 bp in length  
25777  
25876: gap of 100 bp  
25877  
26619: contig of 743 bp in length

26620  
26720  
27445: gap of 100 bp  
26720  
27445: contig of 726 bp in length  
27446  
27545: gap of 100 bp  
27446  
27546: gap of 100 bp  
28290: contig of 745 bp in length  
28291  
28390: gap of 100 bp  
28391  
29130: contig of 740 bp in length  
29131  
29230: gap of 100 bp  
29231  
29955: contig of 726 bp in length  
29956  
30056: gap of 100 bp  
30057  
30806: contig of 750 bp in length  
30807  
30906: gap of 100 bp  
30907  
31648: gap of 100 bp  
31648: contig of 742 bp in length  
31748  
31748: gap of 100 bp  
32469  
32468: contig of 720 bp in length  
32469  
32568: gap of 100 bp  
32568  
33303: contig of 735 bp in length  
33304  
33403: gap of 100 bp  
33304  
34118: contig of 715 bp in length  
34119  
34218: gap of 100 bp  
34219  
34954: contig of 736 bp in length  
34955  
35054: gap of 100 bp  
35055  
35795: contig of 742 bp in length  
35796  
35895: gap of 100 bp  
35797  
35895: gap of 100 bp  
35897  
36640: contig of 744 bp in length  
36641  
37287: gap of 100 bp  
37287: contig of 547 bp in length  
37288  
37387: gap of 100 bp  
37388  
38128: contig of 741 bp in length  
38128  
38228: gap of 100 bp  
38228  
38962: contig of 734 bp in length  
38962  
39062: gap of 100 bp  
39062: contig of 740 bp in length  
39063  
39802: contig of 740 bp in length  
39803  
39902: gap of 100 bp  
39903  
40644: contig of 742 bp in length  
40645  
40744: gap of 100 bp  
40745  
41453: contig of 725 bp in length  
41453  
41563: gap of 100 bp  
41470  
42301: contig of 732 bp in length  
42302  
42401: gap of 100 bp  
42402  
43145: contig of 744 bp in length  
43146  
43245: gap of 100 bp  
43246  
43977: contig of 732 bp in length  
43978  
44077: gap of 100 bp  
44078  
44615: contig of 738 bp in length  
44615  
44815: gap of 100 bp  
44816  
45650: contig of 735 bp in length  
45651  
45750: gap of 100 bp  
45751  
46470: contig of 720 bp in length  
46471  
46570: gap of 100 bp  
46570  
47703: contig of 733 bp in length  
47703  
47403: gap of 100 bp  
47404  
48138: contig of 735 bp in length  
48139  
48238: gap of 100 bp  
48239  
48965: contig of 727 bp in length  
48966  
49065: gap of 100 bp  
49066  
49798: contig of 733 bp in length  
49799  
49898: gap of 100 bp  
49899  
50631: contig of 733 bp in length  
50632  
50731: gap of 100 bp  
50732  
51463: contig of 732 bp in length  
51464  
51563: gap of 100 bp  
51564  
52304: contig of 741 bp in length  
52305  
52404: gap of 100 bp  
52405  
53143: contig of 739 bp in length  
53144  
53243: gap of 100 bp  
53244  
53975: contig of 733 bp in length  
53976  
54075: gap of 100 bp  
54076  
54811: contig of 736 bp in length  
54812  
54911: gap of 100 bp  
54912  
55647: contig of 736 bp in length  
55648  
55747: gap of 100 bp  
55748  
56490: contig of 743 bp in length  
56491  
56590: gap of 100 bp

\* 56591 57342: contig of 752 bp in length  
 \* 57343 57442: gap of 100 bp  
 \* 57443 58192: contig of 750 bp in length  
 \* 58193 58292: gap of 100 bp  
 \* 58293 59017: contig of 725 bp in length  
 \* 59018 59117: gap of 100 bp

Query Match 1.2% Score 22; DB 2; Length 67271;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Ox 974 TAAATTATGATTAAGTTGTTAT 995  
 Db 26933 TAAATTATGATTAAGTTGTTAT 26954

RESULT 10  
 HSDJ863C7 108653 bp DNA linear PRI 08-FEB-2001  
 LOCUS  
 DEFINITION Human DNA sequence from clone RP5-863C7 on chromosome 20p12.3-13. Contains the CSNK2A1 gene for casein kinase 2 alpha 1 polypeptide (EC 2.7.1.37), the 5' end of the gene for a novel protein similar to Drosophila CG17883, ESTs, STSs, GSSs and a Cpg island, complete sequence.

ACCESSION AL049761 GI:12583640  
 VERSION HTG; casein kinase; Cpg island; CSNK2A1.  
 KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens.  
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 108653)  
 AUTHORS Blakey, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire, UK

COMMENT  
 On Jan 27, 2001 this sequence version replaced GI:5738437. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; SW; SWISSPROT; Tr; TrEMBL; Wp; WORMPEP; Information on the WORMPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>

IMPORTANT: This sequence is not the entire insert of clone RP5-863C7 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP5-863C7 is at 1 in this sequence. The true right end of clone RP5-867F11 is at 79232 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP5-863C7 is from the library RPCT-5 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pCYPAC2.  
 Location/Qualifiers  
 1..108653  
 /organism="Homo sapiens"

/mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /map="p12.3-13"  
 /clone="RP5-863C7"  
 /clone1ib="RPCT-5"  
 6..316  
 /note="AluSp repeat: matches 1..312 of consensus"  
 318..602  
 /note="AluO repeat: matches 1..308 of consensus"  
 1396..1694  
 /note="AluO repeat: matches 1..299 of consensus"  
 2730..2848  
 /note="XER31B repeat: matches 353..471 of consensus"  
 3386..3533  
 /note="AluO/FRAM repeat: matches 156..302 of consensus"  
 4442..4742  
 /note="AluO repeat: matches 1..304 of consensus"  
 4786..4965  
 /note="LIMB7 repeat: matches 5955..6160 of consensus"  
 4797..5802  
 /note="Cpg island"  
 /evidence="not\_experimental"  
 5300..5401  
 /note="51 copies 2 mer cc 60% conserved"  
 complement(5417..5612)  
 /gene="d0852M4.2"  
 complement(55417..5612)  
 /gene="d0852M4.2"  
 /product="d0863C7.2 (Novel protein similar to Drosophila CG17883)"  
 /note="match ESTs: Em:BE276163 Em:BE409340 Em:BE276157 Em:BE38289 Em:BE391114 Em:BE391493"  
 /evidence="not\_experimental"  
 complement(5486)  
 /gene="d0852M4.2"  
 /note="continues in Em:AL121747 as d0852M4.2"  
 /codon\_start=1  
 /evidence="not\_experimental"  
 /product="d0863C7.2 (Novel protein similar to Drosophila CG17883)"  
 /protein\_id="CAC28622.1"  
 /db\_xref="GI:12718359"  
 /translation="MALRSAGDGTSGHWDGAEKA"  
 5762..5920  
 /note="WIR repeat: matches 83..251 of consensus"  
 6030..6116  
 /note="WER91C repeat: matches 14..104 of consensus"  
 6141..6375  
 /note="L2 repeat: matches 160..402 of consensus"  
 6378..6453  
 /note="Alu repeat: matches 2..77 of consensus"  
 6779..7055  
 /note="AluO repeat: matches 6..301 of consensus"  
 7184..7480  
 /note="AluX repeat: matches 14..310 of consensus"  
 7829..8090  
 /note="LIMB2 repeat: matches 5769..6032 of consensus"  
 8190..8467  
 /note="WER5B repeat: matches 21..318 of consensus"  
 8503..8792  
 /note="AluSp repeat: matches 1..290 of consensus"  
 8819..9123  
 /note="AluX repeat: matches 1..308 of consensus"  
 9140..9262  
 /note="FRAM C repeat: matches 1..121 of consensus"  
 9268..10101  
 /note="LIMB repeat: matches 1285..2098 of consensus"  
 10102..10400  
 /note="Alu repeat: matches 1..303 of consensus"  
 10401..10623  
 /note="LIMB repeat: matches 1060..1285 of consensus"  
 10760..10942

```

repeat_region /note="AluIo repeat: matches 1. .163 of consensus"
10943. .11411
repeat_region /note="L1R22 repeat: matches 1. .509 of consensus"
11412. .14527
repeat_region /note="HERVK22 repeat: matches 2332. .6837 of consensus"
14526. .14727
repeat_region /note="HERVK22 repeat: matches 699. .905 of consensus"
14716. .15515
repeat_region /note="HERVK22 repeat: matches 1. .808 of consensus"
15516. .15983
misc_feature /note="L1R22 repeat: matches 1. .509 of consensus"
complement(15889. .16375)
/note="match: GSS: Em:AQ475722"
repeat_region /note="AluIo repeat: matches 179. .302 of consensus"
15985. .16108
misc_feature /note="match: GSS: Em:AQ22357"
complement(15993. .16331)
16081. .16361
misc_feature /note="match: STS: Em:G37195"
17097. .17135
repeat_region /note="L2 repeat: matches 2155. .2194 of consensus"
17136. .17414
repeat_region /note="AluIo repeat: matches 3. .300 of consensus"
17415. .17615
repeat_region /note="L2 repeat: matches 1964. .2155 of consensus"
17693. .18019
repeat_region /note="HALL repeat: matches 1194. .1528 of consensus"
18258. .18558
repeat_region /note="AluX repeat: matches 1. .299 of consensus"
19341. .19638
repeat_region /note="AluX repeat: matches 2. .299 of consensus"
19649. .19951
repeat_region /note="AluIo repeat: matches 1. .305 of consensus"
19982. .20269
repeat_region /note="AluIo repeat: matches 1. .290 of consensus"
20290. .20519
repeat_region /note="AluIo repeat: matches 59. .286 of consensus"
20791. .20953
repeat_region /note="AluIo repeat: matches 1. .168 of consensus"
20964. .21003
repeat_region /note="L2 repeat: matches 2627. .2701 of consensus"
21179. .21251
repeat_region /note="L2 repeat: matches 2627. .2701 of consensus"
21728. .22028
repeat_region /note="AluIo repeat: matches 1. .304 of consensus"
22211. .22325
repeat_region /note="MER91B repeat: matches 12. .123 of consensus"
22863. .23076
repeat_region /note="MER46A repeat: matches 1. .235 of consensus"
23316. .23359
repeat_region /note="L2 copies 2 mer tg 93% conserved"
24196. .24443
misc_feature /note="match: STS: Em:G27906"
24895. .25060
misc_feature /note="match: GSS: Em:A2244707"
25073. .25276
misc_feature /note="match: GSS: Em:A2244707"
complement(25221. .25784)
/note="match: GSS: Em:AQ508904"
/note="match: GSS: Em:AQ568163"
complement(25777. .86902)
/note="match: GSS: Em:AQ568163"
complement(25777. .86902)
Gene /Gene="CSNR2A1"
complement(join(25777. .27157, 29457. .29543, 30508. .30656,
31759. .31859, 32861. .32962, 35335. .35445, 36800. .36883,
40802. .40861, 42336. .42386, 42914. .43015, 48159. .48310,
51532. .51741, 86753. .86902))
/rName="CSNR2A1"
/product="dJ863C7.1.1 (casein kinase 2, alpha 1
polypeptide (EC 2.7.1.37))"
/note="isoform 1
match: cDNAs: Em:X54962 Em:J02853 Em:M55265 Em:U51866

```

```

Query Match 1.2% Score 22; DB 9; Length 108653;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 229 AACCTGAATCCATCATTTT 250
Db 9244 AACTGAATCCATCATTTT 9223

RESULT 11
AC121962
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Locus: MUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.edu
Project Information
Center project name: M_B8025105
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phred; version 0.990319
Consensus quality: 152181 bases at least Q40
Consensus quality: 152299 bases at least Q30
Consensus quality: 152299 bases at least Q20
Insert size: 175000; agarose-fp
Insert size: 155087; sum-of-contigs
Quality coverage: 13.31 in Q20 bases; agarose-fp
Quality coverage: 11.42 in Q20 bases; sum-of-contigs

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 14531: contig of 14531 bp in length
* 14532 14631: gap of unknown length
* 14632 152537: contig of 137906 bp in length
* 152538 152637: gap of unknown length
* 152638 153354: contig of 717 bp in length
* 153355 153454: gap of unknown length

```



FEATURES \* 153537: contig of 83 bp in length.  
Location/Qualifiers  
source 1. 153537  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10090"  
/chromosome="X"  
/clone="RP24-25515"  
1. 14531  
misc\_feature /note="assembly\_name:Contig52"  
14632. 152537  
/note="assembly\_name:Contig53"  
152638. 153354  
misc\_feature /note="assembly\_name:Contig19"  
153455. 153537  
/note="assembly\_name:Contig4"  
BASE COUNT 41644 a 34816 c 36006 g 40771 t 300 others  
ORIGIN

Query Match 1.2%; Score 22; DB 2; Length 15337;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1560 AAAAGTGTGATGAAATGCT 1581  
|||||  
99134 AAAAGTGTGATGAAATGCT 99155

RESULT 12  
LOCUS AC040953 160536 bp DNA linear HTG 24-AUG-2002  
DEFINITION Homo sapiens chromosome 20 clone RP11-147L21 map 20, WORKING DRAFT  
SEQUENCE AC040953  
AC040953.2 GI:7960448  
HTG: HTGS PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 160536)  
Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome 20, clone RP11-147L21  
Unpublished  
2 (bases 1 to 160536)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Basilen,V., Bede,F.,  
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,  
Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., Deavellano,K., Dewar,K., Diaz,J.S.,  
Dodgson,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagsos,B., Heatford,A., Horton,L.,  
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McSheeters,R.,  
Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliveira,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B.,  
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
COMMENT

Submitted (11-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 160536)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Basilen,V., Bede,F.,  
Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G.,  
Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., Deavellano,K., Dewar,K., Diaz,J.S.,  
Dodgson,S., Domingo,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hagsos,B., Heatford,A., Horton,L.,  
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Larocque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,  
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McSheeters,R.,  
Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neil,D., Oliveira,T.M., Oliver,J., Peterson,K., Pierre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severi,P., Spencer,B.,  
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.

Submitted (12-APR-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On May 21, 2000 this sequence version replaced gi:1534154.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
Project Information  
Center project name: L9154  
Center clone name: 147 L 21

Summary Statistics  
Sequencing vector: M13; M77815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 147817 bases at least Q40  
Consensus quality: 153894 bases at least Q30  
Consensus quality: 156395 bases at least Q20  
Insert size: 162000; agarose-fp  
Insert size: 157736; sum-of-contigs  
Quality coverage: 4.3 in Q20 bases; agarose-fp  
Quality coverage: 4.4 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
consists of 29 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1  
1332: contig of 1332 bp in length  
1432: gap of 100 bp  
1433  
3055: contig of 1623 bp in length  
3056  
3155: gap of 100 bp  
3156  
4227: contig of 1072 bp in length  
4228  
4327: gap of 100 bp  
4328  
6154: contig of 1827 bp in length  
6155  
6254: gap of 100 bp  
6255  
8472: contig of 2218 bp in length  
8473  
8572: gap of 100 bp  
8573  
1191: contig of 2619 bp in length  
1192  
1192: gap of 100 bp  
1199: contig of 3099 bp in length  
1199  
14390: contig of 3099 bp in length  
14391  
14490: gap of 100 bp  
14491  
17075: contig of 2585 bp in length  
17076  
17175: gap of 100 bp  
17176  
19838: contig of 2663 bp in length  
19839  
19938: gap of 100 bp  
19939  
22383: contig of 2445 bp in length  
22384  
22483: gap of 100 bp

```

* 22484 25382: contig of 2899 bp in length
* 25383 25482: gap of 100 bp
* 25483 28079: contig of 2597 bp in length
* 28080 28179: gap of 100 bp
* 28180 32156: contig of 3977 bp in length
* 32157 32256: gap of 100 bp
* 32257 35644: contig of 3388 bp in length
* 35645 35744: gap of 100 bp
* 35745 39479: contig of 3735 bp in length
* 39480 43753: gap of 100 bp
* 43754 43853: contig of 4174 bp in length
* 43854 50636: contig of 6783 bp in length
* 50637 50736: gap of 100 bp
* 50737 57916: contig of 7180 bp in length
* 57917 58016: gap of 100 bp
* 58017 63882: contig of 5866 bp in length
* 63883 63982: gap of 100 bp
* 63983 71647: contig of 7665 bp in length
* 71648 71747: gap of 100 bp
* 71748 78978: contig of 7231 bp in length
* 78979 79078: gap of 100 bp
* 79079 84661: contig of 5583 bp in length
* 84662 84761: gap of 100 bp
* 84762 90578: contig of 5817 bp in length
* 90579 90678: gap of 100 bp
* 90679 98969: contig of 8291 bp in length
* 98970 99069: gap of 100 bp
* 99070 111241: contig of 12172 bp in length
* 111242 111341: gap of 100 bp
* 111342 121097: contig of 9756 bp in length
* 121098 121197: gap of 100 bp
* 121198 130885: contig of 9688 bp in length
* 130886 130985: gap of 100 bp
* 130986 143583: contig of 12598 bp in length
* 143584 143684: gap of 100 bp
* 143684 160336: contig of 16853 bp in length.

```

## FEATURES

```

source
  1. 160536
    /organism="Homo sapiens"
    /mol_type="genomic DNA"
    /db_xref="taxon:9606"
    /chromosome="20"
    /map="20"
    /clone="RP11-147121"
    /clone_1db="RP11-1 Human Male BAC"
  1. 1332
    /note="assembly_fragment"
  1433. 3055
    /note="assembly_fragment"
  3156. 4227
    /note="assembly_fragment"
  4328. 6154
    /note="assembly_fragment"
  6235. 8472
    /note="assembly_fragment"
  8573. 11191
    /note="assembly_fragment"
  11292. 14390
    /note="assembly_fragment"
  14491. 17075
    /note="assembly_fragment"
  17176. 19638
    /note="assembly_fragment"
  19339. 22383
    /note="assembly_fragment"
  22484. 25382
    /note="assembly_fragment"
  25483. 28079
    /note="assembly_fragment"
  28180. 32156
    /note="assembly_fragment"
  32257. 35644
    /note="assembly_fragment"

```

```

misc_feature /note="assembly_fragment"
              35745..39479
              /note="assembly_fragment"
misc_feature 39580..43753
              /note="assembly_fragment"
misc_feature 43854..50636
              /note="assembly_fragment"
misc_feature 50737..57916
              /note="assembly_fragment"
misc_feature 58017..63882
              /note="assembly_fragment"

```

```

Query Match 1.2% Score 22; DB 2; Length 160536;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 229 AACTGGAATCCATCATTTT 250
Db 111004 AACTGGAATCCATCATTTT 111025

```

## RESULT 13

## AC025089

## LOCUS

## DEFINITION

## AC025089

## ACCESSION

## VERSION

## KEYWORDS

## SOURCE

## ORGANISM

## REFERENCE

## AUTHORS

## TITLE

## JOURNAL

## REFERENCE

## AUTHORS

```

1 (bases 1 to 165870)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, U., Barna, N., Bastien, V., Beda, F.,
Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G.,
Campoliano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S.,
Collamore, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Gajagan, J., Gardina, S., Ginde, S., Goyette, N., Graham, L.,
Grand-Pierre, N., Grant, G., Hages, B., Heatford, A., Horton, L.,
Holland, J.C., Iliev, I., Johnson, R., Jones, C., Kahn, J., Karatas, A.,
Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehocque, J.,
Levine, R., Liu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPherson, R.,
Meidrim, J., Menus, L., Mihova, T., Miranda, C., Mianga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Tagimas, J.,
Teftaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Yael, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (04-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 165870)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Boukhgalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J.,
Chazaro, B., Choquel, Y., Colangelo, M., Collins, S., Collamore, A.,
Cooke, A., Cooke, P., Dearrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,
Galagan, J., Gardina, S., Ginde, S., Goid, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Hages, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,
Lamazares, R., Landers, T., Lehocque, J., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., MacDonald, P., Major, D., Marquis, N.,

```

TITLE  
JOURNAL  
COMMENT

Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Neldrum, J.,  
Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C.,  
Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, J., Peterson, R., Phunkhang, P., Pierre, N.,  
Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,  
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schaner, S.,  
Schuback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Strauss, N., Sudramanah, A., Talamas, J., Tesfaye, S.,  
Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,  
Vassiliev, H., Vei, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission

Submitted (11-JUN-2002) Whitehead Institute/MIT Center for Genome  
Research, 330 Charles Street, Cambridge, MA 02141, USA  
On Jun 11, 2002 this sequence version replaced gi:17998733.

All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WtBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L7911

Center clone name: 275\_B\_19

# FEATURES

## source

Location/Qualifiers  
1. 165870  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-275B19"  
/clone\_lib="RPC1-11 Human Male BAC"  
2. 506  
/rpt\_family="L1"  
510. 714  
/rpt\_family="THB1C"  
717. 790  
/rpt\_family="L1"  
792. 2114  
/rpt\_family="WSTA-int"  
2115. 2520  
/rpt\_family="WSTA"  
2544. 2726  
/rpt\_family="LIMB8"  
2767. 4003  
/rpt\_family="LIMB8"  
complement(4004..4372)  
/rpt\_family="THB1B"  
complement(4373..4838)  
/rpt\_family="THB1B-int"  
complement(4836..5274)  
/rpt\_family="THB1B-int"  
complement(5277..5640)  
/rpt\_family="THB1B"  
5642. 6034  
/rpt\_family="LIMB8"  
6035. 6285  
/rpt\_family="ALUy"  
6286. 6413  
/rpt\_family="LIMB8"  
complement(6436..6798)  
/rpt\_family="LIMB5"  
8239. 8274  
/rpt\_family="AT-rich"  
10207. 10241  
/rpt\_family="(TA)n"  
10786. 10807  
/rpt\_family="AT-rich"  
11181. 11206

/rpt\_family="(TA)n"  
complement(11656..11884)  
/rpt\_family="LIM2"  
complement(12975..13256)  
/rpt\_family="L2"  
complement(13897..13957)  
/rpt\_family="LIP2"  
complement(20587..21007)  
/rpt\_family="MSTA"  
21161. 21189  
/rpt\_family="(TATC)n"  
21228. 21403  
/rpt\_family="MIR"  
complement(21668..21768)  
/rpt\_family="MSTA"  
22088. 22268  
/rpt\_family="MIR"  
complement(22451..23151)  
/rpt\_family="LIM2"  
complement(23890..23958)  
/rpt\_family="MIR"  
complement(23984..24330)  
/rpt\_family="TIG9er1"  
complement(24331..24448)  
/rpt\_family="FLAY\_A"  
complement(24449..26551)  
/rpt\_family="TIG9er1"  
complement(28565..29190)  
/rpt\_family="HERV16"  
29241. 29262  
/rpt\_family="AT-rich"  
29335. 29398  
/rpt\_family="(TCTA)n"  
30674. 30696  
/rpt\_family="AT-rich"  
30747. 30800  
/rpt\_family="(CA)n"  
31017. 31037  
/rpt\_family="AT-rich"  
31061. 31088  
/rpt\_family="AT-rich"  
complement(31950..32252)  
/rpt\_family="ALUy"  
32597. 33487  
/rpt\_family="MER4D"  
34612. 34646  
/rpt\_family="AT-rich"  
complement(34653..35005)  
/rpt\_family="MER4A"  
complement(35006..35100)  
/rpt\_family="MER4A"  
complement(35105..35306)  
/rpt\_family="LIM4A"  
35460. 35491  
/rpt\_family="(TAAA)n"  
complement(36326..36519)  
/rpt\_family="LIM4"  
37109. 37114  
/note="<30 qual SNGL region"  
complement(37161..37376)  
/rpt\_family="MLTIL"  
complement(37407..40118)  
/rpt\_family="Char163"  
complement(40124..40415)  
/rpt\_family="MLTIL"  
40528. 41815  
/rpt\_family="LIM4A"  
41814. 41848  
/rpt\_family="AT-rich"  
41988. 42299  
/rpt\_family="ALUSg"  
43159  
/note="probably G, possibly C"

unsure 43194, 43195  
/note="probably CC"  
43830, 44208  
repeat\_region /rpt\_family="LRR33"

Query Match 1.2%; Score 22; DB 9; Length 165870;  
Best Local Similarity 100.0%; Pred No.12;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

974 TAAATTATGATTAAGTTGTAT 995

Db 141362 TAAATTATGATTAAGTTGTAT 141383

RESULT 14 AC009336 175667 bp DNA linear PRI 17-OCT-2000  
LOCUS Homo sapiens chromosome 2, clone RP11-387A1, complete sequence.  
DEFINITION AC009336  
VERSION AC009336.13 GI:10835353

KEYWORDS HTG.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens

REFERENCE 1 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
1 (bases 1 to 175667)  
Birren, B., Linton, L., Nusbaum, C. and Lander, E.

JOURNAL 2 Unpublished  
REFERENCE 2 (bases 1 to 175667)  
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Becker, R., Benn, J., Brown, A.,

Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Darrellano, K., Depayre, E., Devon, K., Dewar, K., Donegan, J., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kam, L., Karatas, A., Lehoczy, D., Lien, C., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P., Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Tesfaye, S., Tornuella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (16-AUG-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 175667)

AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bede, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Darrellano, K., Dewar, K., Dodge, S., Domingo, M., Doyle, M., Fenebor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J., Landers, T., Largocque, K., Lehoczy, J., Levine, R., Liu, G., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., Menga, V., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Peterson, K., Pletre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (17-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE 4 (bases 1 to 175667)  
AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Barna, N., Bastien, V., Bede, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Darrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Heaford, A., Horton, L., Iliev, I., Johnson, R., Jones, C., Kam, L., Karatas, A., Largocque, K., Lamazeres, P., Landers, T., Lehoczy, J., Levine, R., Liu, G., MacDonald, P., Marquis, N., McCarthy, M., McEwan, P., Menga, V., McEwan, P., Morrow, J., Menus, L., Minova, T., Morrow, J., Murphy, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pletre, N., Pisan, C., Pollara, V., Raymond, C., Riback, M., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Sougnaz, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission  
JOURNAL Submitted (17-OCT-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Oct 17, 2000 this sequence version replaced gi:6584494. All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WTHR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

Project Information  
Center project name: L935  
Center clone name: 387\_A\_1

FEATURES

source

Location/Qualifiers  
1..175667  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-387A1"  
/clone\_lib="RP11-387A1 Human Male BAC"  
complement(11..243)  
/rpt\_family="L1MCc"  
complement(250..561)  
/rpt\_family="Alusg"  
complement(600..816)  
/rpt\_family="L1MCc"  
1457..1638  
/rpt\_family="Alusx"  
complement(2510..2632)  
/rpt\_family="L2"  
complement(2638..3056)  
/rpt\_family="MLT1J2"  
3197..3503  
/rpt\_family="LTR16A1"  
3602..3722  
/rpt\_family="LRR2B"  
3728..3875  
/rpt\_family="Aluy"  
3881..4171  
/rpt\_family="LRR2B"  
4986..5275  
/rpt\_family="AluCb"  
5278..5326  
/rpt\_family="CA)n"  
5818..5944  
/rpt\_family="L2"  
5946..5986

```

repeat_region /rpt_family="AT_rich"
6023. .6350
repeat_region /rpt_family="LTR37A"
6527. .6845
repeat_region /rpt_family="AlubB"
7059. 7110
repeat_region /rpt_family="T-rich"
8666. .8888
repeat_region /rpt_family="AT_rich"
complement(12825. .12945)
repeat_region /rpt_family="FLAM A"
complement(13097. .13390)
repeat_region /rpt_family="AlubB"
13738. .13773
repeat_region /rpt_family="AT_rich"
14420. .14720
repeat_region /rpt_family="AlusX"
14848. .14898
repeat_region /rpt_family="AT_rich"
14901. .15236
repeat_region /rpt_family="AlusX"
15431. .15657
repeat_region /rpt_family="MIR"
17847. .17869
repeat_region /rpt_family="AT_rich"
complement(18152. .18447)
repeat_region /rpt_family="AlubB"
19551. .19767
repeat_region /rpt_family="MIR"
complement(19815. .19851)
repeat_region /rpt_family="MIR"
complement(20616. .20913)
repeat_region /rpt_family="AlusX"
21282. .21310
repeat_region /rpt_family="AT_rich"
complement(21790. .21811)
repeat_region /rpt_family="L2"
complement(21812. .22126)
repeat_region /rpt_family="AlubB"
complement(22127. .22259)
repeat_region /rpt_family="L2"
22260. .22594
repeat_region /rpt_family="MER44D"
complement(22595. .22736)
repeat_region /rpt_family="L2"
22821. .23239
repeat_region /rpt_family="ML1K"
23634. .23790
repeat_region /rpt_family="L2"
complement(24009. .24368)
repeat_region /rpt_family="MER46B"
24965. .25287
repeat_region /rpt_family="L1MC4a"
25382. .25989
repeat_region /rpt_family="L1MCc"
25990. .26293
repeat_region /rpt_family="AlubB"
26294. .26954
repeat_region /rpt_family="L1MCc"
26955. .27310
repeat_region /rpt_family="THB1B"
27311. .27439
repeat_region /rpt_family="L1MCc"
complement(27440. .27636)
repeat_region /rpt_family="AlusG/x"
27665. .27927
repeat_region /rpt_family="AlusG/x"
27932. .28395
repeat_region /rpt_family="L1MCc"
complement(28396. .28708)

```

Query Match 1.2%; Score 22; DB 9; Length 175667;  
 Best Local Similarity 100.0%; Pred. No. 12;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1037 TGAATATATGTTTCTAGCTTA 1058  
 |||||  
 Db 8376 TGAATATATGTTTCTAGCTTA 8355

RESULT 15  
 AC100799  
 LOCUS  
 DEFINITION  
 AC100799 Homo sapiens chromosome 8, clone RP11-1082F24, complete sequence.  
 VERSION  
 AC100799.2 GI:19551148  
 KEYWORDS  
 HTG.  
 SOURCE  
 ORGANISM  
 Homo sapiens (human)  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 2 (bases 1 to 194149)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens chromosome 8, clone RP11-1082F24  
 Unpublished

1 (bases 1 to 194149)  
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
 Homo sapiens chromosome 8, clone RP11-1082F24  
 Unpublished  
 2 (bases 1 to 194149)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barr, N., Bastien, V., Boguslavsky, L., Bouknight, B.,  
 Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
 Choquet, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
 Cooke, P., Dearlano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
 Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hages, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,  
 Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G.,  
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., McSheeters, R., Meldrum, J.,  
 Menus, L., Mihova, T., Miesha, V., Murphy, T., Naylor, J., Nguyen, C.,  
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rose, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
 JOURNAL  
 REFERENCE  
 AUTHORS  
 Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 194149)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barr, N., Bastien, V., Bloom, T., Boguslavsky, L.,  
 Bouknight, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
 Chazaro, B., Choquet, Y., Colangelo, M., Collins, S., Collymore, A.,  
 Cook, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
 Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hages, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R.,  
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
 Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,  
 Mihova, T., Miesha, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
 Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rose, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (20-MAR-2002) Whitehead Institute/MIT Center for Genome

## COMMENT

Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Mar 20, 2002 this sequence version replaced gi:17048169.  
 All repeats were identified using RepeatMasker:  
 Smith, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## ----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIRB

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center Project name: L20718

Center clone name: 1082\_F\_24

## FEATURES

## Source

## Location/Qualifiers

```

1..194148
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-1082F24"
/clone_id="RP11-Human Male BXC"
complement(679..1326)
/rpt_family="L1MB1"
1327..1353
/rpt_family="(TTTG)n"
complement(1354..1772)
/rpt_family="L1MB1"
1773..7802
/rpt_family="L1PA2"
7803..7938
/rpt_family="L1MB1"
7939..8167
/rpt_family="L1PA8"
8168..8292
/rpt_family="L1MB1"
8310..8424
/rpt_family="T-rich"
complement(8453..8861)
/rpt_family="L1MB1"
8862..10050
complement(8886..10050)
/rpt_family="L1MCA"
10063..10763
complement(10063..11034)
/rpt_family="L1MCA"
11035..12038
complement(11019..12038)
/rpt_family="L1PA8"
12039..12343
/rpt_family="AluSx"
complement(12344..14680)
/rpt_family="L1PA8"
14681..14936
complement(14681..14936)
/rpt_family="AluSc"
14937..17408
complement(14937..17408)
/rpt_family="L1PA8"
17409..17898
complement(17409..18229)
/rpt_family="L1MCA"
18230..18436
complement(18234..18436)
/rpt_family="L1MCA"
18437..18470
/rpt_family="(TTTGG)n"
complement(18471..18757)
/rpt_family="L1MCA"
18758..19605
complement(18759..19605)
/rpt_family="L1MCA"
19606..21065
complement(19594..21065)
/rpt_family="L1M1"
21066..21482
complement(21480..21482)
/rpt_family="MSTD"

```

```

repeat_region complement(121887..21964)
/rpt_family="L1M1"
21965..22270
/rpt_family="AluY"
repeat_region complement(22271..22349)
/rpt_family="L1M1"
22350..23495
complement(22501..23495)
/rpt_family="MER1A"
23496..24427
/rpt_family="AT-rich"
complement(24503..26314)
/rpt_family="AluSc"
26315..26415
/rpt_family="AT-rich"
26416..26613
/rpt_family="(TA)n"
complement(27317..27633)
/rpt_family="T2"
27634..29647
complement(29358..29647)
/rpt_family="AluSc"
30303..30326
/rpt_family="AT-rich"
30327..30795
/rpt_family="AT-rich"
30796..31204
complement(30988..31204)
/rpt_family="MR"
31205..31381
complement(31303..31381)
/rpt_family="MER63"
31382..31893
complement(31808..31893)
/rpt_family="L1M"
31894..32171
/rpt_family="L1MB1"
32172..32469
/rpt_family="AluSc"
32470..32750
/rpt_family="L1MB1"
32751..33831
/rpt_family="AT-rich"
33832..34323
/rpt_family="(TAA)n"
34324..34913
complement(34382..34913)
/rpt_family="L1PA4"
34914..35046
/rpt_family="L1P2"
35047..36168
/rpt_family="MSTB"
36169..37766
/rpt_family="MSTB-int"
37767..38185
/rpt_family="MSTB"
38186..39083
complement(38519..39083)
/rpt_family="MLT2B3"
39084..39873
/rpt_family="AT-rich"
39874..41682
/rpt_family="(TA)n"
41683..42591
complement(42591..42848)
/rpt_family="MERS7-int"
42849..43932
complement(42849..43932)
/rpt_family="L1PA"
43933..44217
/rpt_family="(TA)n"
44218..44443
complement(44443..44443)

```

Query Match 1.2% Score 22; DB 9; Length 194149;

Best Local Similarity 100.0%; Pred. No. 12; Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 974 TAAATTATGATTAAGTTGTAT 995

DB 30654 TAAATTATGATTAAGTTGTAT 30675

Wed Nov '26 09:11:03 2003

us-09-831-083-1.rge

Page 17

Search completed: November 25, 2003, 18:16:37  
Job time : 4401 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 25, 2003, 13:13:48 ; Search time 2616 Seconds

(without alignments)  
16565.321 Million cell updates/sec

Title: US-09-831-083-1

Perfect score: 1783

Sequence: 1 atccaactctgcatcttga.....tagagcgatcaagctgaacc 1783

Scoring table: IDENTITY NUC

Gapop 10.0 ; Gapext 1.0

Searched: 22781392 segs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_estbun:\*  
3: em\_estin:\*  
4: em\_estm:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gss\_hum:\*  
18: em\_gss\_iv:\*  
19: em\_gss\_pln:\*  
20: em\_gss\_vit:\*  
21: em\_gss\_fun:\*  
22: em\_gss\_mam:\*  
23: em\_gss\_mus:\*  
24: em\_gss\_pro:\*  
25: em\_gss\_rtd:\*  
26: em\_gss\_png:\*  
27: em\_gss\_vrl:\*  
28: gb\_gss1:\*  
29: gb\_gss2:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	114	6.4	1101	29	CNS00EO7
2	104.6	5.9	1101	29	CNS00EOVL
3	103.6	5.8	1200	13	BX415878
4	102	5.7	1200	13	BX436510

Result No.	Score	Query Match	Length	DB ID	Description
5	101	5.7	1101	29	CNS00EO7
6	101	5.7	1200	13	BX415878
7	99.4	5.6	928	29	CNS00DKY
8	99	5.6	1101	29	CNS00EOVL
9	98.2	5.5	1201	9	AL565455
10	98.2	5.5	1201	13	BX458623
11	97	5.4	1201	9	AL536104
12	97	5.4	1201	13	BX443774
13	96.6	5.4	1101	29	CNS0021T
14	95.8	5.4	1201	9	AL536104
15	95.8	5.4	1201	13	BX335216
16	95.2	5.3	1201	13	BX458623
17	95	5.3	1200	13	BX437758
18	94.8	5.3	661	29	CNS020VT
19	94.8	5.3	1098	13	BX377526
20	94	5.3	1201	13	BX395109
21	94	5.3	1201	13	BX461824
22	93.4	5.2	1200	13	BX436510
23	92.8	5.2	1092	29	CNS020K7
24	92.6	5.2	1277	29	CC253231
25	92.4	5.2	1056	13	BX415058
26	92.2	5.2	1201	13	BX443774
27	91.8	5.1	1098	13	BX377526
28	91.2	5.1	1101	29	CNS0039G
29	91.2	5.1	1201	13	BX366070
30	90.8	5.1	1101	29	CNS0039G
31	90.8	5.1	1201	13	BX462207
32	90.6	5.1	1101	29	CNS0021T
33	90.4	5.1	994	29	CNS04NOU
34	90.2	5.1	854	9	AL522840
35	90	5.0	1101	29	CNS003BD
36	89	5.0	1094	29	CNS012FZ
37	88.6	5.0	1148	9	AL527799
38	88.6	5.0	1165	13	BX383869
39	88.6	5.0	1201	13	BX381083
40	88.4	5.0	1201	13	BX426629
41	88	4.9	1147	14	CD387645
42	87.8	4.9	1201	13	BX366070
43	87.6	4.9	836	29	CNS01100
44	87.2	4.9	1056	13	BX415058
45	86.8	4.9	1200	29	CNS016CO

## ALIGNMENTS

RESULT 1  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

CNS00EO7 1101 bp DNA linear GSS 04-JUN-1999  
Drosophila melanogaster genome survey sequence TERT end of BAC:  
BACR29P01 of RPCT-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL069440  
AL069440.1 GI:4949583  
GSS.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Submission  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuhiro Oseogawa and  
Aaron Mammeter in Pieter de Jong's laboratory in the Department of

Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

Location/Qualifiers

1..1101

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone\_id="RPCI-98"

/note="end : TET3"

## BASE COUNT

366 a 66 c 104 g 351 t 214 others

## ORIGIN

Query Match 6.4%; Score 114; DB 29; Length 1101;  
Best Local Similarity 39.3%; Pred. No. 4,5e-05;

Matches 254; Conservative 98; Mismatches 291; Indels 4; Gaps 2;

QY 812 AATCTACAAAGAGAGATTTCACCTAGAGAGTATGGAGTTAAAGACACA 871  
DB AAGAGGAAAAAACAATKAAARGGGGAATAATWTKGAAAGATATAAAAA 468  
QY 872 CATTAGGGGAGGTGTTAAATTATGTTGTTGACCCACCTACCTTGTAGCTTA 931  
DB AAT 528  
QY 932 TAAGAAATTTGTAATCATCATTAATTAATGCTTATTTAAATTTATGATTAAGTT 991  
DB AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 529  
QY 992 GTATCATTTAGA---TTGAGAAACCAATATGCTGCTGTTTGTATATGTT 1048  
DB TTTWATTTAT 583  
QY 1049 TTTCTATGTTACTTTCTTCAAGCTATATATATATATATATATATATAT 1108  
DB TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 648  
QY 1109 AAAAAAATGTTAATGAATTCATAGATATGATATTCATCAAGTCCAAATCCATCA 1168  
DB AAAAAATTTAT 708  
QY 1169 TAGAAATTTAGTACAAAACGTAACTCAAAAATATCTCTATTTTAAATTTTCAACAAT 1228  
DB TAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 828  
QY 1229 ATATATAA-ATATCTGTTATTTTAAATTTTAAATTAATTAATTTTCACTGACCTTT 1287  
DB TAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 828  
QY 1289 AAGATACCAACCAATATATATATAGATATTTTATCTTAAATTTTGAATCTCT 1347  
DB AATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 948  
QY 1348 CAATATATCTGATATTTTATATATATTTGTCATTTTCTATATGTTTATAGCTTACC 1407  
DB AATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1008  
QY 1408 CTTATATCTTGGTCAAACTAGTAAATTCATATATATAGTTGTGAAGG 1454  
DB AAAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1008  
QY 1009 AAAAAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1055

## RESULT 2

## CNS00EVL

LOCUS 1101 bp DNA linear GSS 04-JUN-1999

DEFINITION Drosophila melanogaster genome survey sequence 17 end of BAC:

BAC29B23 of RPCI-98 library from Drosophila melanogaster (fruit

fly), genomic survey sequence.

ACCESSION AL0659706

VERSION AL0659706.1 GI:4949849

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

- Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr)

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).

The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosawa and Aaron Mammotser in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

## FEATURES

Location/Qualifiers

1..1101

/organism="Drosophila melanogaster"

/mol\_type="genomic DNA"

/db\_xref="taxon:7227"

/clone\_id="BACR29B23"

/note="end : T7"

## BASE COUNT

419 a 91 c 60 g 299 t 232 others

## ORIGIN

Query Match 5.9%; Score 104.6; DB 29; Length 1101;  
Best Local Similarity 34.8%; Pred. No. 0.00066;  
Matches 218; Conservative 120; Mismatches 285; Indels 4; Gaps 1;

QY 830 ATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 889  
DB AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 889  
QY 890 AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 949  
DB AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 949  
QY 949 AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 1009  
DB AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 1009  
QY 1009 AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 1069  
DB AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 1069  
QY 1069 AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 1129  
DB AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 1129  
QY 1129 AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 1189  
DB AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 1189  
QY 1189 AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 1245  
DB AATTTTAACTTGAAGAGATGAGAGTTAAAGCAACATTAAGGGGAGGTGA 1245

```

FEATURES
  source
    Location/Qualifiers
      1..1200
        /organism="Homo sapiens"
        /mol_type="RNA"
        /db_xref="taxon:9606"
        /clone="CSOCAF002YK09"
        /tissue_type="THYMUS"
        /clone_id="Homo sapiens THYMUS"

```

/note=vector: pcmsport 6: 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pcmsport 6 vector. Library was not normalized.

BASE COUNT 422 a 116 c 144 g 411 t 107 others

Query Match 5.7%; Score 102; DB 13; Length 1200;  
Best Local Similarity 41.5%; Pred. No. 0.0013;  
Matches 217; Conservative 56; Mismatches 250; Indels 0; Gaps 0;

810 AAAATACACAAAGAGAGATTTTAAACACTAGAGAGATGAGAGATTAAGACCA 869  
576 AAATTA 635  
870 CACATTAAGGGGAGTGAATTAATGATGTTGTAACCACTACCTTAGTAGTAT 929  
636 ATAAAAATTTTAAAAAATTTAAAAAATTTTAAAAAATTTTAAAAAATTTT 695  
930 TATAGAAATTTGATGATCATCATTAATTAATGCTTATTTAAATTAATGATAAG 989  
696 TTTATATATATTTTATATATATATATATATATATATATATATATATATAT 755  
990 TTGATCATTAAGATGAGAAACCAATAGCTCTGCTGATTTTGAATTAATGTT 1049  
756 ATTATATATATTTTAAAAAATTAATTAATTTTAAAAAATTTTAAAAAATTT 815  
1050 TCTATGTTACTTTCTTCAACCTATATATACTTTGATGATGATGATGATGGA 1109  
816 TTTTATTTTATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAAT 875  
1110 AAAAAATGTAAGATTCATAGAAATTTGATTTCAAGTCCAAATCCATCAT 1169  
876 AAAAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 935  
1170 AGAATTTAGTACAAAGCTACCAAAATTTCTTATTTAAATTTTACACATA 1229  
936 TTTAAAAATTTATATTTTATTTTAAAAAATTTTAAAAAATTTTAAAAAAT 995  
1230 TAAAAATTTCTTATTTTAAATTTTACATTAATTAATTAATTAATTAATTA 1289  
996 TTTATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAA 1055  
QY 1290 AATACCAACCAATTAATTAATTAATTAATTTATTTATTTATTTAT 1332  
Db 1056 AAAAAATTTTATTAATTAATTAATTTTAAATTTTAAATTTTAAATTT 1098

RESULT 5  
CNS00E07/c 1101 bp DNA linear GSS 04-JUN-1999

LOCUS Drosophila melanogaster genome survey sequence TBT3 end of BAC:  
BACR29P01 of Rpci-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL069440.1 GI:4949583

KEYWORDS GSS.

SOURCE Drosophila melanogaster (fruit fly)

ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)

AUTHORS Genoscope.

TITLE Direct Submission

JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 Evry cedex - FRANCE (E-mail : seque@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information

please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kanuoyo Osoegwa and Aaron Mammeter in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named Rpci-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at [http://bacpac.med.buffalo.edu/drosophila\\_bac.htm](http://bacpac.med.buffalo.edu/drosophila_bac.htm).

#### FEATURES

source  
1. 1101  
/organism="Drosophila melanogaster"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:7227"  
/clone="BACR29P01"  
/clone\_11b="Rpci-98"  
/note="end : TBT3"

BASE COUNT 366 a 66 c 104 g 351 t 214 others

Query Match 5.7%; Score 101; DB 29; Length 1101;  
Best Local Similarity 38.0%; Pred. No. 0.0018;  
Matches 206; Conservative 80; Mismatches 256; Indels 0; Gaps 0;

906 ACCACACCTACCTTAGTATATATAGAAATTTGATATCATCATTAATTAATG 965  
1019 AAAAAATTTTAAATTTTATTTTAAATATATATATTTTAAATATATATTTTAA 960  
QY 966 TCTTATTTAAATTAATGATTAAGTGTATCTTAAAGTTGAGAAACCAATAGCTC 1025  
Db 959 TACATATATTTTAAAAATTTTAAATTAATTAATTAATTAATTAATTAATTA 900  
QY 1026 GTCTGATTTTGAATTAATGTTTCTATGTTTCTCAAGCTATATTAATTAAT 1085  
Db 899 TTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 840  
QY 1086 TGTAAATGCTAAATTTGATGCGGAAAAAATGCTGATGAATTCATGAAATTA 1145  
Db 839 TAAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 780  
QY 1146 TTCAAGTCCAAATTCATGATTAATTAATTAATTAATTAATTAATTAATTA 1205  
Db 779 TAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 720  
QY 1206 CTTATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1265  
Db 719 TTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 660  
QY 1266 TTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1325  
Db 659 AATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 600  
QY 1326 TCTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1385  
Db 599 AAAAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 540  
QY 1386 TTTCTTATGTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 1445  
Db 539 TTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAATTT 480  
QY 1446 TT 1447  
Db 479 WT 478

RESULT 6  
BX415878/c 1200 bp mRNA linear EST 15-MAY-2003  
LOCUS BX415878 Homo sapiens THYMUS Homo sapiens cDNA clone CS00AP008Y104  
DEFINITION 5-PRIME, mRNA sequence.  
ACCESSION BX415878

```

VERSION      EX15878.1  GI:30765550
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE    1 (bases 1 to 1200)
              L1, N.B., Gruber, C., Jesse, J. and Polayes, D.
              Full-length cDNA libraries and normalization
              JOURNAL
              Unpublished
              Contact: Genoscope
              Genoscope - Centre National de Sequencage
              BP 191 91006 Evry cedex - France
              Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr
              Library was constructed by Life Technologies, a division of
              Invitrogen. Contact: Feng Liang Email: fliang@life.com URL:
              http://fulllength.invitrogen.com/Invitrogen Corporation 1600
              Faraday Avenue Genoscope sequence ID: CS0CAP008BE02CP1.
              Location/Qualifiers
                1. 1200
                   /organism="Homo sapiens"
                   /mol_type="mRNA"
                   /db_xref="taxon:9606"
                   /clone="CS0CAP008Y104"
                   /class_type="THYMUS"
                   /clone_lib="Homo sapiens THYMUS"
                   /note=Vector: pCMVSPORT 6; 1st strand cDNA was primed
                   with a NotI-oligo(dT) primer. Five prime end enriched,
                   double-strand cDNA was digested with Not I and cloned into
                   the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                   Library was not normalized."
BASE COUNT   428 a      70 c      85 g      447 t      170 others
ORIGIN
Query Match 5.7%; Score 101; DB 13; Length 1200;
Best Local Similarity 37.8%; Pred. No. 0.0018;
Matches 273; Conservative 102; Mismatches 335; Indels 13; Gaps 2;
OY          832 TTTAACAAGCTTAGAGAGTAAAGAGTAAAGAGACATTAAGGGAGTGTAA 891
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          1200 TTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1141
OY          892 ATTAATGCTGTAAACACACCTCTAGTAAAGTATTAAGAAAATTGTAATCA 951
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          1140 TTTATATTTTATATTAATAAATAAATAAATAAATAAATAAATAAATA 1081
OY          952 CATTAATATTAATGCTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1011
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          1080 TTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1021
OY          1012 ACCAAATAGCTGCTGTGAATTTTGAATTAATGTTTCTAGTAACTTCTCAAGC 1071
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          1020 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 971
OY          1072 CTATATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1131
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          970 TTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 911
OY          1132 TAGAATATATGATATTTCAAGTCCAAATCATCATAGAAATTTAGTCAAAAGCTAA 1191
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          910 TTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 851
OY          1192 CTCAAAATATCTCTTATTTTAAATTTTAAACAATTAATAAATTTCTTATTTAA 1251
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          850 TTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 791
OY          1252 ATTTACATATATATATTTTACCTGCTACCTTAGATACACACCAATATTAATA 1311
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          790 AATATTTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 731
OY          1312 CTGAGATATTTATCTTAATATTTGAGATCTCTCAATATATATGAT--ATTATTT 1368
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          730 TTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 671

```

```

OY          1369 TATATTTGTCATATTTCTTATGTTTAAAGTTAAACCTTATATCTGTCAACTAG 1428
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          670 TTAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 611
OY          1429 TATATATATATGATTTTGAAGACATTTGACATCTTGAAACATTTGTTTAACT 1488
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          610 TTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 551
OY          1489 TGTGGAATTTAAAGCTATTAATAACATTTGACATCTTATTAATATCTTCC 1548
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          550 AAAATATTAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 491
OY          1549 TTT 1551
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          490 TTT 488
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 7
CNS00DX     928 bp   DNA   linear   GSS 04-JUN-1999
LOCUS       Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION  BACR27A24 of RPCL-98 library from Drosophila melanogaster (fruit
ACCESSION   AL071865
VERSION     AL071865.1  GI:4948170
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM    Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
              Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
              Ephydroidea; Drosophilidae; Drosophila.
              1 (bases 1 to 928)
REFERENCE   1
  AUTHORS   Direct Submission
  TITLE     Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage:
  JOURNAL   BP 191 91006 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)
  COMMENT   - Web: www.genoscope.cns.fr
             Determination of this BAC-end sequence was carried out as part of a
             collaboration with the Berkeley Drosophila Genome Project (BDGP).
             The BDGP is constructing a physical map of the Drosophila
             melanogaster genome using these BACs. For further information
             please see http://www.fruitfly.org The BDGP Drosophila
             melanogaster BAC library was prepared by Kazuo Osoegawa and
             Aaron Mammose in Peter de Jong's laboratory in the Department of
             Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
             NY. The library is named RPCL-98 and was constructed by partial
             EcoRI digestion of Drosophila DNA provided by the BDGP from the
             isogenic strain y2; cn bw sp, the same strain used for the library
             and EST libraries. A more detailed description of the library
             and how to order individual BAC clones, the entire library, or
             filters for hybridization from the BACPAC Resource Center can be
             found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
             Location/Qualifiers
               1. 928
                  /organism="Drosophila melanogaster"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:7227"
                  /clone="BACR27A24"
                  /clone_lib="RPCL-98"
                  /note="end: T7"
BASE COUNT   262 a      70 c      84 g      321 t      191 others
ORIGIN
Query Match 5.6%; Score 99.4; DB 29; Length 928;
Best Local Similarity 35.2%; Pred. No. 0.003;
Matches 151; Conservative 97; Mismatches 181; Indels 0; Gaps 0;
OY          969 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1028
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB          479 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 538
OY          1029 TTGATTTTGAATTTGTTTCTATAGTTACTTTCTCAAGCTTATTAATAAATA 1088
           |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```



http://fulllength.invitrogen.com/InvitrogenCorporation1600  
Faraday Avenue Genoscope sequence ID : CS0DF005BH09NP1.  
Location/Qualifiers

FEATURES  
SOURCE  
1..1201  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DF005Y018"  
/issue\_type="FETAL BRAIN"  
/dev\_stage="fetal"  
/clone\_lib="Homo sapiens FETAL BRAIN"  
/note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end  
enriched, double-strand cDNA was digested with Not I and  
cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
vector. Library was not normalized."

BASE COUNT 550 a 82 c 71 g 313 t 185 others  
ORIGIN

Query Match 5.5%; Score 98.2; DB 9; Length 1201;  
Best Local Similarity 32.3%; Pred. No. 0.004;  
Matches 241; Conservative 134; Mismatches 372; Indels 0; Gaps 0;

```
QY 832 TTATACAACTTAGAGAGTAATGGAGTTAAGACACATTACGGGAGGTGTTAA 891
DB 1196 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 1137
QY 892 ATTATGCTGTGTAACACACACTCTTGTAGTATTATAGAAATGTATCATCA 951
DB 1136 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 1077
QY 952 CATTAATATTATGCTCTTATTTAAATATGATTAAGTTGATTAAGTTGAAA 1011
DB 1076 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 1017
QY 1012 ACCAAATAGTCCTGCTGATTTGTGATTTGATTTGTTGCTTCTGCTCAAGC 1071
DB 1016 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 957
QY 1072 CTATATATAAACTTTGTAATCTTAATGTATGCTGAAAAAAATGTATGATTC 1131
DB 956 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 897
QY 1132 TAGAAATTATGCTATTTCAAGTCCAAATCCATCATAGAAATTTGTCAAAAGCT 1191
DB 896 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 837
QY 1192 CTCAAAATATCTCTTATTTTAAATTTTACACAAATATAAAATCTCTTATTT 1251
DB 836 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 777
QY 1252 ATTTTACATATATTAATTTATCCCTGTCACTTTTGAATACCAACAATTTA 1311
DB 776 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 717
QY 1312 CTGATATTTTATTTCTTAATTTTGAATCTCTCAATATATCTATTTATTT 1371
DB 716 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 657
QY 1372 ATTTGTGATATTTTCTTATGTTTGAAGTTAACCTTATATCTGTCACAACT 1431
DB 656 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 597
QY 1432 TTCAATATATAGTGTGAGAGCAACATTGACATCTTGAACATTTGTTTAACT 1491
DB 596 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 537
QY 1492 TGAATATGTAAGGTAATATAACATTCAGATTTGACCATCTTATATCTCTCT 1551
DB 536 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 477
QY 1552 GTCTTTTAAATAAGTGATGAAAT 1578
```

DB 476 TWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTATWTAT 450

RESULT 10  
BX458623/c 1201 bp mRNA linear EST 22-May-2003  
LOCUS BX458623 Homo sapiens PLACENTA Homo sapiens cDNA clone CS0DE006YG03  
DEFINITION 5-PRIME, mRNA sequence.  
ACCESSION BX458623  
VERSION BX458623.1 GI:31019361  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
1 (bases 1 to 1201)  
Li, W. B., Gruber, C., Jessee, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 9800.f For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0DE006AD020P1&cluster=9800.f. Contact :  
Feng Liang Email : fliang@lifech.com URL :  
http://fulllength.invitrogen.com/InvitrogenCorporation1600  
Faraday Avenue Genoscope sequence ID : CS0DE006AD02QPL.

## FEATURES

## SOURCE

1..1201  
Location/Qualifiers

/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DE006YG03"  
/issue\_type="PLACENTA"  
/clone\_lib="Homo sapiens PLACENTA"  
/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
with a NotI-oligo(dT) primer. Five prime end enriched,  
double-strand cDNA was digested with Not I and cloned into  
the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
Library was not normalized."

BASE COUNT 407 a 174 c 161 g 355 t 104 others  
ORIGIN

Query Match 5.5%; Score 98.2; DB 13; Length 1201;  
Best Local Similarity 42.9%; Pred. No. 0.004;  
Matches 222; Conservative 54; Mismatches 226; Indels 2; Gaps 1;

```
QY 924 AAGTATTAATAGAAATTTGTAATCATCATTAATTAATTTGCTTATTTAAATT 983
DB 1040 AATKATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 981
QY 984 ATTAAGTTGATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1043
DB 980 TTTAATTTATATKATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 921
QY 1044 TTGTTTTCATGTTTCTTCTTCAAGCCTATATAAAAATTTGTAATGCTAAT 1103
DB 920 TTTAATAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 861
QY 1104 GCTGGAATAAATATGTAATGTAATCAATGAAATTAATGTAATCAAGTCAAT 1163
DB 860 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 801
QY 1164 ATCAATGAAATTTAGTACAAACGTAATCAATCAATTAATTTCTTATTTAAAT 1223
DB 800 TTTAATTAATTAATTTTATTTTAAATTAATTAATTAATTAATTAATTTAAAT 743
QY 1224 ACAATTAATAATTTCTTATTTTAAATTTTCAATTAATTAATTAATTAATTA 1283
```



Db 742 AAATAATTTTATTTATTTTAAAGTATAAAAAAATTTTATTTT 683  
 Qy 1284 CTTAGATACCACCAATATTAATCTAGATTTTATCTTATATTTGAT 1343  
 Db 682 TTTAAATTAATAAAAAAATTTATTTTAAATTTTATTTTATTTT 623  
 Qy 1344 CTCTCAATATATCGAATTTTATTTTATTTTGTGATTTTCTTATCTTTAGATT 1403  
 Db 622 TTTTAAATTAADAMTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 563  
 Qy 1404 AACCTTATATCTT 1417  
 Db 562 WMTTATTTTATTTT 549

RESULT 11  
 AL536104 1201 bp mRNA linear EST 31-MAY-2003  
 LOCUS AL536104 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone  
 DEFINITION CSODF022YC18 5-PRIME, mRNA sequence.  
 ACCESSION AL536104  
 VERSION AL536104.2 GI:31260974  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization.  
 JOURNAL Unpublished  
 COMMENT On Feb 13, 2001 this sequence version replaced gi:12799597.  
 CONTACT: Genoscope - Centre National de Sequencage  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CSODF022B09QPL.  
 Location/Qualifiers  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODF022YC18"  
 /tissue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_lib="Homo sapiens FETAL BRAIN"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-strand cDNA was digested with Not I and  
 cloned into the Not I and EcoRV sites of the pCMVSPORT 6  
 vector. Library was not normalized."  
 BASE COUNT 359 a 119 c 146 g 409 t 168 others  
 ORIGIN

Query Match 5.4%; Score 97; DB 9; Length 1201;  
 Best Local Similarity 34.4%; Pred. No. 0.0056;  
 Matches 158; Conservative 100; Mismatches 201; Indels 0; Gaps 0;

Qy 923 TAAGATTAAGAAATTTGATCATCATTAATTAATTTGCTTATTTAAATTAAT 982  
 Db 732 TTTTAAATTAATTAATTTTATTTTAAATTTTATTTTATTTTATTTTATTTT 791  
 Qy 983 GATAAGCTTATGATTAAGTTGAGAAACCAATAGCTCTGATTTTGAATTT 1042  
 Db 792 WMTTATTTTATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTT 851  
 Qy 1043 ATTGTTTCTATGTTACTTTCTTCAAGCTATATAAACTTTGTAATGCTAATTTGA 1102  
 Db 852 WMTTATTTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 911

Qy 1103 TCGTGGAAAAAATGCTGATGATTCATGAAATTTGCTATTTCCAAAGTCCAAATC 1162  
 Db 912 TATWAAATTTTATTTATTTATTTAGTATWAAWAAWAAATATATATATWAAWAAWAAW 971  
 Qy 1163 CATCATTAAGAAATTTAGTACCAAAAGCTAATCTCAAAAATATCTCTATTTAAATTTAC 1222  
 Db 972 AAAAAATATATTAATAAAAAAATWAAATATTTATTTATTTTATTTTATTTTATTTT 1031  
 Qy 1222 AACCAATATAAAAAATTTCTTATTTTAAATTTTACAAATATATATTTATTCACCTGTCA 1282  
 Db 1032 ATWATAAAAAAATWATTTTATTTATTTATTTATTTTAAWAAWAAWAAWAAWAAWAAW 1091  
 Qy 1283 CCTTTAGAAATACCACCAATATTTAATCTAGATTTTATTTTCTTATTAATTTTGA 1342  
 Db 1092 WMTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1151  
 Qy 1343 TCTCTCAATATATCGAATTTTATTTTATTTTGTGCTA 1381  
 Db 1152 DMTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 1190

RESULT 12  
 BX443774 1201 bp mRNA linear EST 15-MAY-2003  
 LOCUS BX443774 Homo sapiens T CELLS (JURKAT CELL LINE) Homo sapiens cDNA  
 DEFINITION clone CSODH007YF12 5-PRIME, mRNA sequence.  
 ACCESSION BX443774  
 VERSION BX443774.1 GI:30772178  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT Contact: Genoscope - Centre National de Sequencage  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of  
 Invitrogen. This sequence belongs to sequence cluster 7885.f for  
 more information about this cluster, see  
 http://www.genoscope.cns.fr/  
 cgi-bin/cluster.cgi?seq=CSODH007DC06P1;cluster=7885.f. Contact :  
 Feng Liang Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com/Invitrogen Corporation 1600  
 Faraday Avenue Genoscope sequence ID : CSODH007DC06QPL.  
 Location/Qualifiers  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODH007YF12"  
 /tissue\_type="T CELLS (JURKAT CELL LINE)"  
 /cell\_line="JURKAT CELL LINE"  
 /clone\_lib="Homo sapiens T CELLS (JURKAT CELL LINE)"  
 /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed  
 with a NotI-oligo(dT) primer. Five prime end enriched,  
 double-strand cDNA was digested with Not I and cloned into  
 the Not I and EcoRV sites of the pCMVSPORT 6 vector.  
 Library was not normalized."  
 BASE COUNT 394 a 158 c 153 g 395 t 101 others  
 ORIGIN

Query Match 5.4%; Score 97; DB 13; Length 1201;  
 Best Local Similarity 38.6%; Pred. No. 0.0056;  
 Matches 262; Conservative 72; Mismatches 345; Indels 0; Gaps 0;

Qy 952 CATTAATTAATTTGCTTATTTAAATTAATGATTAAGCTTATCATTTAGATTGAGAA 1011  
 Db 515 CTTTGTGCTAAACCTTTTATTAACATGCTCAATATAAAAGCTGACCTCAAAAAAAA 574



```

QY 1012 ACCAATAGTCCTGCTGATTTTGAATTATGTTTCTAGTGTACTTTCTTCAAGC 1071
DB 575 AAAAAAAAAAAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 634
QY 1072 CTATATATAAACTTGTAAGCTAAATGTATGCTGAAAAAATGTAATGATTCGA 1131
DB 635 TTTTTTTTTTTTTTTTTTTTTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTT 694
QY 1132 TAGAATATATGATTTTCAAGTCAAAATCAATCAATAGAAATTTAGTCAAAAGTAA 1191
DB 695 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATATAA 754
QY 1192 CTCAAAATATCTCTATTTTAAATTTACACATATATAAATATCTCTATTTTAA 1251
DB 755 AAAAAAAAAADTTTATATATATATATATATATATATATATATATATATATAT 814
QY 1252 ATTTACATATATATATATATATATATATATATATATATATATATATATAT 1311
DB 815 TTTTATATATAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 874
QY 1312 CTATGATTTTATTTCTTAATAATTTGACATCTTCAATATATCTGATATTTAT 1371
DB 875 TTTTATATAAATAATTTTTTTTTTTTTTTTTTTTTTTTATATATATATATATAT 934
QY 1372 ATTTGCTCATATTTCTTATGTTTGAAGTTAACCTTATCTTGTCATCTAGTAA 1431
DB 935 AATATATATATATATATATATATATATATATATATATATATATATATATAT 994
QY 1432 TTCAATATATAGTTTGTAGAGACATGACATCTTGAACATTTGTTTAACTTGA 1491
DB 995 ATATATATATAAATAATATATATATATATATATATATATATATATATATAT 1054
QY 1492 TGAATCTTAAAGTATATAAATCATGAGATATGACATCTATATATATCTCTT 1551
DB 1055 ATTTTATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1114
QY 1552 GTCTTTTAAATAAGTGCATGAAATGCTCATGTTAGCTAGAGTCTGCTGGCT 1611
DB 1115 TTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1174
QY 1612 GTGATATCTCAATTCATT 1630
DB 1175 AATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 1193

```

RESULT 13  
CNS0021J/c 1101 bp DNA linear GSS 03-JUN-1999  
Drosophila melanogaster genome survey sequence TET3 end of BAC #  
BACR05N11 of RPCI-98 library from Drosophila melanogaster (fruit  
fly), genomic survey sequence.  
AL061936  
AL061936.1 GI:4940214  
GSS.  
Drosophila melanogaster (fruit fly)  
Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
Phytophagae; Drosophilidae; Drosophila.  
1 (bases 1 to 1101)  
Genoscope.  
Direct Substition  
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
Determination of this BAC-end sequence was carried out as part of a  
collaboration with the Berkeley Drosophila Genome Project (BDGP).  
The BDGP is constructing a physical map of the Drosophila  
melanogaster genome using these BACs. For further information  
please see <http://www.fruitfly.org> The BDGP Drosophila  
melanogaster BAC library was prepared by Kazuo Oosawa and  
Aaron Mamoser in Pieter de Jong's laboratory in the Department of

```

FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone_1ib="RPCI-98"
/note="end : TET3"
location/Qualifiers
BASE COUNT 631 a 7 c 28 g 289 t 146 others
ORIGIN
Query Match 5.4%; Score 96.6; DB 29; Length 1101;
Best Local Similarity 42.6%; Pred. No. 0.0064;
Matches 258; Conservative 37; Mismatches 310; Indels 0; Gaps 0;
QY 954 TATATATATGCTCTTATTTAAATATGATTAAGTTGTATCATTAAGTGAAGAAC 1013
DB 916 TTAACCTTAAATAAATATATATATATATATATATATATATATATATATATAT 857
QY 1014 CAATATAGTCTGCTGCTGATTTTGAATATGTTTCTATGTTACTTTCTCAAGCT 1073
DB 856 AAAAAAAAAAATATATATATATATATATATATATATATATATATATATATAT 797
QY 1074 ATATATAAACTTGTAAGCTAAATGTATGCTGAAAAAATGTAATGATTCATA 1133
DB 796 AAAAAATTAATATAAATAATTTTATATATATATATATATATATATATATAT 737
QY 1134 GAATATATGATTTTCAAGTCAAAATCCATCATATAAATTTAGTCAAAAGTAACT 1193
DB 736 AAAAAATTAATATAAATAATTTTATATATATATATATATATATATATATAT 677
QY 1194 CAATATATGCTCTTATTTTAAATTTTCAACATATATAAATAATCTTATTTTAA 1253
DB 676 AAAAAATTAATATAAATAATTTTATATATATATATATATATATATATATAT 617
QY 1254 TTTACATATATATATATATATATATATATATATATATATATATATATAT 1313
DB 616 TTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 557
QY 1314 TAGATATTTTATTTCTTAATAATTTTGAAGTCTCAATATATCTGATTTTATAT 1373
DB 556 AAAAAAAAAAAAAAAAAATTTTATATAAATAAATAAATAAATAAATAAATAAATA 497
QY 1374 TTGCTCATTTTCTTATGTTTGAAGTTAACCTTATATCTTGTCATCTGTAAT 1433
DB 496 TTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 437
QY 1434 CAATATATGATTTTGTAGAGACATGACATCTGAAACATGTTTAACTTGTG 1493
DB 436 TTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 377
QY 1494 GATGTTTAAAGTATATAAATCATGAGATATGACATCTATATATATCTCTTGT 1553
DB 376 TTTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 317
QY 1554 CTATT 1558
DB 316 TTTT 312

```

RESULT 14  
AL536104/c 1201 bp mRNA linear EST 31-MAY-2003  
LOCUS AL536104 Homo sapiens FETYL BRAIN Homo sapiens cDNA clone  
DEFINITION CSDF022Yc18 5-PRIME, mRNA sequence.

ACCESSION AL536104  
 VERSION GI:31260974  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT On Feb 13, 2001 this sequence version replaced gi:1279597.  
 CONTACT: Genoscope - Centre National de Sequencage  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSODF022B8092P1.

FEATURES  
 source  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSODF022XC18"  
 /tissue\_type="FETAL BRAIN"  
 /dev\_stage="fetal"  
 /clone\_lib="Homo sapiens FETAL BRAIN"  
 /note="Organ: Brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo (dt) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was not normalized."

BASE COUNT  
 ORIGIN 359 a 119 c 146 g 409 t 168 others

Query Match 5.4%; Score 95.8; DB 9; Length 1201;  
 Best Local Similarity 35.4%; Pred. No. 0.0078;  
 Matches 207; Conservative 107; Mismatches 270; Indels 1; Gaps 1;

803 TTTTCTAAATATACACAAAGAGAGAGATTTCACCACTTGAGAGAGATGGAGTTAA 862  
 1201 WWHYHTATNTTAT 1142  
 863 AGAGCAACACATTAAAGGAGAGAGTTAAATTAATGAGTGAACCACTACCTTAG 922  
 1141 AATAATATCAAMWMTATWHTATTTTATTTTATTTTATTTTATTTTATTTTAT 1082  
 923 TAAGATATATAGAAATTTGTAATCATCATCATTAATTAATGCTTATTTAAATTA 982  
 1081 WAMAYMTWMTTATTAATATATATATATATATATATATATATATATATATATAT 1022  
 983 GATTAAGTTGATCTTAAAGATTGAGAAACCAATAGTCTGCTTGAATTTGAATT 1042  
 1021 WAAAAAATAATATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTT 962  
 1043 ATTGTT-TTCATGTTACTTTCTTCAAGCTATATATATATATATATATATATAT 1101  
 961 ATWMTAT 902  
 1102 ATCTGCGAAAAATGCTATATGAATTCATGAATTAATGATTTCAAGTCCAAAT 1161  
 901 AATAAAAT 842  
 1162 CCATCAATATAGAAATTTGATCAAAACGTAACCTCAAAATATATCTATATTTT 1221  
 841 WTTTAT 782  
 1222 CAACATATATATATATATCTTTATTTTATTTTATTTTATTTTATTTTATCT 1281  
 781 WYAT 722

1282 ACCCTTAGATACACACATATATATATATATATATATATATATATATATATAT 1341  
 721 TATATTTAT 662

1342 ATCTCAAT 1386  
 661 WMAAT 617

RESULT 15  
 EX35216/c  
 LOCUS EX35216/c  
 DEFINITION Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA  
 ACCESSION EX35216  
 VERSION EX35216.1 GI:30341402  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1201)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished  
 COMMENT Contact: Genoscope - Centre National de Sequencage  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
 Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 714.f for more information about this cluster, see  
 http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CSOD1012BD050P1&cluster=714.f. Contact : Feng Liang Email : fliang@lifetech.com url : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CSOD1012BD050P1.

FEATURES  
 source  
 1..1201  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="CSOD1012XC10"  
 /tissue\_type="PLACENTA COT 25-NORMALIZED"  
 /clone\_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"  
 /note="1st strand cDNA was primed with a NotI-oligo (dt) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoRV sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT  
 ORIGIN 391 a 115 c 137 g 453 t 105 others

Query Match 5.4%; Score 95.8; DB 13; Length 1201;  
 Best Local Similarity 43.7%; Pred. No. 0.0078;  
 Matches 255; Conservative 46; Mismatches 271; Indels 12; Gaps 2;

887 TTAAT 946  
 1164 TAAAT 1105  
 947 CATCAAT 1006  
 1104 TTAT 1045  
 1007 AGAAATCAAT 1066  
 1044 WTAAT 985  
 1067 CAAGCTAT 1117  
 984 AATAAT 925  
 1118 TGTAT 1177



This Page Blank (uspto)